**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| (51) International Patent Classification 6 : C12N 15/60, 9/88, 1/19, 1/21, 15/81, 15/70, C12P 7/24, 7/40, 7/04 | A3 | (11) International Publication Number: **WO 99/58648**<br>(43) International Publication Date: 18 November 1999 (18.11.99) |
|---|---|---|

(21) International Application Number: PCT/IB99/00807

(22) International Filing Date: 5 May 1999 (05.05.99)

(30) Priority Data:
09/078,173 13 May 1998 (13.05.98) US

(71) Applicants: FIRMENICH SA [CH/CH]; 1, route des Jeunes, P.O. Box 239, CH–1211 Geneva 8 (CH). VANDERBILT UNIVERSITY [US/US]; 305 Kirkland Hall, Nashville, TN 37240 (US). THE UNIVERSITY OF ZURICH [CH/CH]; Zollikerstrasse 107, CH–8008 Zurich (CH).

(72) Inventors: WHITEHEAD, Ian, Michael; 19 Almond Street, Singapore 677861 (SG). SLUSARENKO, Alan, John; Schievenhövel 52, B–4728 Hergenrath (BE). WASPI, Urs; Erlachstrasse 25, CH–8003 Zurich (CH). GASKIN, Duncan, James, Horatio; 3 Linden Cottages, Whitchurch Hill, Reading RG8 7PH (GB). BRASH, Alan, Richard; 1105 Holly Tree Farms, Brentwood, TN 37027 (US). TIJET, Nathalie; 3920 Westlawn Drive, Nashville, TN 37209 (US).

(74) Agent: SALVATERRA–GARCIA, Maria de Lurdes; Firmenich S.A., P.O. Box 239, CH–1211 Geneva 8 (CH).

(81) Designated States: BR, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

**Published**
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(88) Date of publication of the international search report:
24 February 2000 (24.02.00)

(54) Title: GUAVA (*PSIDIUM GUAJAVA*) 13–HYDROPEROXIDE LYASE AND USES THEREOF

(57) Abstract

The present invention relates to fatty acid 13–hydroperoxide lyase protein from guava (*Psidium guajava*) and the gene encoding the protein. Expression systems for recombinant guava 13–hydroperoxide lyase and methods of using recombinant guava 13–hydroperoxide lyase for the production of green notes are provided.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

# INTERNATIONAL SEARCH REPORT

Int ional Application No
PCT/IB 99/00807

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/60 C12N9/88 C12N1/19 C12N1/21 C12N15/81
C12N15/70 C12P7/24 C12P7/40 C12P7/04

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | EP 0 801 133 A (GIVAUDAN-ROURE (INTERNATIONAL) S.A.) 15 October 1997 (1997-10-15) cited in the application page 2, line 3 -page 5, line 51; examples 1-5 | 1-51 |
| Y | US 5 464 761 A (FIRMENICH S.A.) 7 November 1995 (1995-11-07) cited in the application column 2, line 1 -column 2, line 65; figure 1; examples 1,2 | 1-51 |

-/—

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 16 December 1999 | 11/01/2000 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL – 2280 HV Rijswijk<br>Tel. (+31–70) 340–2040, Tx. 31 651 epo nl,<br>Fax: (+31–70) 340–3016 | Donath, C |

1

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 99/00807

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication,where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | MATSUI, K. ET AL.: "Bell pepper fruit fatty acid hydroperoxide lyase is a cytochrome P450 (CYP74B)" FEBS LETTERS, vol. 394, no. 1, 23 September 1996 (1996-09-23), pages 21-24, XP002126014 cited in the application the whole document | 1-51 |
| A | FAUCONNIER, M.-L. ET AL.: "Purification and characterization of tomato leaf (Lycopersicon esculentum Mill.) hydroperoxide lyase" JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY, vol. 45, no. 11, November 1997 (1997-11), pages 4232-4236, XP002126015 cited in the application the whole document | 1-51 |
| A | SHIBATA, Y. ET AL.: "Purification and properties of fatty acid hydroperoxide lyase from green bell pepper fruits" PLANT AND CELL PHYSIOLOGY, vol. 36, no. 1, 1995, pages 147-156, XP002126016 cited in the application the whole document | 1-51 |
| A | MATSUI, K. ET AL.: "Fatty acid hydroperoxide cleaving enzyme, hydroperoxide lyase, from tea leaves" PHYTOCHEMISTRY, vol. 30, no. 7, 1991, pages 2109-2113, XP002126017 cited in the application the whole document | 1-51 |

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/IB 99/00807

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| EP 0801133 A | 15-10-1997 | JP 10033185 A | 10-02-1998 |
| US 5464761 A | 07-11-1995 | BR 9305533 A | 27-12-1994 |
| | | DE 69320302 D | 17-09-1998 |
| | | DE 69320302 T | 28-01-1999 |
| | | WO 9324644 A | 09-12-1993 |
| | | EP 0597069 A | 18-05-1994 |
| | | JP 7500252 T | 12-01-1995 |

Express Mail No. EL055857549US

ATTORNEY DOCKET NO.: 06027.0001

Utility Patent

# APPLICATION

# FOR

# UNITED STATES LETTERS PATENT

## ON

## Guava (*Psidium guajava*)13-Hydroperoxide Lyase and Uses Thereof

## BY

**Ian M. Whitehead, a citizen of Great Britain; Alan Slusarenko, a citizen of Great Britain; Urs Wäspi, a citizen of Switzerland; Duncan Gaskins, a citizen of Great Britain; Alan Brash, a citizen of Great Britain; Nathalie Tijet, a citizen of France, residing, respectively, at 323 Rue de Bernex, Bernex, Geneva, CH-1233, Switzerland; Schievenhövel 52, B-4728, Hergenrath, Belgium; Erlachstrasse 25, Zürich 8003, Switzerland; 3 Linden Cottages, Whitchurch Hill, Reading RG 87 PH, England; 1105 Holly Tree Farms, Brentwood, Tennessee 37027; 3920 Westlawn Drive, Nashville, Tennessee 37209.**

# Guava (*Psidium guajava*) 13-Hydroperoxide Lyase and Uses Thereof

## BACKGROUND OF THE INVENTION

### Field of the Invention

The present invention relates to fatty acid 13-hydroperoxide lyase protein from guava (*Psidium guajava*) and the gene encoding the protein. The present invention also relates to the means for expressing guava 13-hydroperoxide lyase and methods of using guava 13-hydroperoxide lyase in the field of organic synthesis.

### Background Art

Green notes, which include n-hexanal, hexan-1-ol, 2(*E*)-hexen-1-al, 2(*E*)-hexen-1-ol and 3(Z)- hexen-1-ol (also known as pipol), are used widely in flavors, particularly fruit flavors, to impart a fresh green character. Furthermore, green notes are essential for fruit aroma and are used extensively in the aroma industry. The demand for natural green notes has grown to exceed their supply from traditional sources such as mint (*Mentha arvensis*) oil. This has motivated research efforts toward finding alternative natural ways of obtaining these materials.

The synthesis of green note compounds starts from free (polyunsaturated) fatty acids such as linoleic (9(Z), 12(Z)-octadecadienoic) and α-linolenic (9(Z), 12(Z), 15(Z)-octadecatrienoic) acids. In nature, these acids are released from cell membranes by lipolytic enzymes after cell damage. Fatty acid 13-hydroperoxides are formed by the action of a specific lipoxygenase (13-LOX) and are subsequently cleaved by a specific 13-hydroperoxide lyase (13-HPOL) into a C-6 aldehyde and a C-12 ω-oxoacid moiety. The aldehydes can subsequently undergo thermal isomerization and/or be reduced by dehydrogenase enzymes to give the other C-6 products (i.e., green notes) mentioned above (Hatanaka, 1993; Hatanaka, et al., 1987).

The enzyme 13-HPOL has proven difficult to study because it is membrane bound and is present in only small quantities in plant tissue. It was identified for the first time in banana fruits (Tressl and Drawert, 1973) and was subsequently studied in a number of different plant materials, including watermelon seedlings (Vick and Zimmerman, 1976), apple and tomato fruits (Schreier and Lorenz, 1982), tomato leaves (Fauconnier et al., 1997), cucumber seedlings (Matsui, et al, 1989), and soybean seedlings (Olias et al., 1990). The enzyme has been purified to apparent homogeneity from tea leaves (Matsui et al., 1991) and, more recently, from green bell pepper fruits (Shibata et al., 1995), tomato leaves (Fauconnier et al., 1997), and banana (European Patent Application, Publication No. EP 0801133 A2). The various characteristics of 13-HPOLs that have been studied are summarized in Table 1.

**Table 1 - Summary of the Properties of 13-HPOL from Different Sources**

| Enzyme Source | Native Mass (kD) | Sub-Unit Structure | Structure | pH Optimum | pI |
|---|---|---|---|---|---|
| Cucumber | - | - | - | 8.0 | - |
| Green pepper | 170 | 55 | Trimer | - | - |
| Soybean seedlings | 240-260 | 62 | Tetramer | 6.0 - 7.0 | - |
| Tea leaves | - | 53 and 55 | - | 7.5 | - |
| Tomato fruits | 200 | - | - | 5.5 | 5.8 - 6.1 |
| Watermelon | >250 | - | - | 6.0 - 6.5 | - |
| Tomato leaves | 216 | 73 | Trimer | 7.0 | 4.9 |

Guava has recently been identified as an excellent source of freeze-stable 13-HPOL for use in this synthetic pathway. Guava 13-HPOL is currently used in an industrial process for the production of green notes (U.S. Pat. No. 5,464,761). In this process, a solution of the required 13-hydroperoxides is made from linoleic or linolenic

acid (obtained from sunflower and linseed oils, respectively) using freshly prepared soybean flour as a source of 13-LOX. This solution is then mixed with a freshly prepared puree of whole guava (*Psidium guajava*) fruit, as the source for 13-HPOL. The aldehyde products are then isolated by distillation. When the alcohols are required, fresh baker's yeast is added to the hydroperoxide solution before it is mixed with the guava puree. This yeast contains an active alcohol dehydrogenase enzyme that reduces the aldehydes as they are formed by 13-HPOL.

There are a number of disadvantages to this industrial process. The principal disadvantage is the requirement of large quantities of fresh guava fruit. Such a requirement means that the process has to be operated in a country where fresh guava fruit is cheaply and freely available. Even when such a site is found, availability is limited to the growing season of the fruit. Good quality guava fruit, for example, is only available for ten months of the year in Brazil.

A second disadvantage is that the desired enzyme activities are rather dilute in the sources employed. This means that relatively large amounts of soy flour (5%), guava puree (41%) and yeast (22%) have to be used in the process. The large volumes of these crude materials that are required for industrial production place physical constraints on the yields of green notes that can be achieved.

A third disadvantage is that it is a large-volume batch process, which, by its nature, does not make maximum use of the 13-HPOL enzyme's catalytic activity, is relatively labor intensive and generates a large amount of residual organic material. The residual organic material must subsequently be transported to a compost farm or otherwise discarded.

The present invention overcomes these limitations and disadvantages related to the source of guava 13-HPOL by providing purified and recombinant guava 13-HPOL proteins, nucleic acids, expression systems, and methods of use thereof.

## SUMMARY OF THE INVENTION

The present invention provides a fatty acid 13-hydroperoxide lyase (13-HPOL) and a nucleic acid encoding the lyase. In particular, it provides a guava-derived protein having 13-hydroperoxide lyase function and a nucleic acid encoding such protein. The present invention further provides a nucleic acid which specifically hybridizes with the nucleic acid encoding guava 13-hydroperoxide lyase under stringent conditions and which does not hybridize at the same stringent conditions to the nucleic acid encoding green pepper or banana 13-hydroperoxide lyase.

The present invention also provides means for expressing recombinant 13-hydroperoxide lyase. Specifically, a vector for the expression of a guava 13-hydroperoxide lyase comprising the nucleic acid of the present invention and cells containing the exogenous nucleic acid of the present invention are provided. Also provided is a method of expressing the recombinant protein produced by the transformed cells comprising optimizing active lyase function of the recombinant protein.

The present invention further provides methods of using recombinant 13-hydroperoxide lyase. Specifically, the present invention provides a method of cleaving a 13-hydroperoxide of linoleic acid into a n-hexanal and a $C_{12}$-oxocarboxylic acid. Also provided is a method of preparing n-hexanal, 3-(Z)-hexen-1-al, 2-(E)-hexen-1-al, or their corresponding alcohols from 13-hydroperoxy-octadeca-9,11-dienoic acid or 13 hydroperoxy-octadeca-9,11,15-trienoic acid.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the alignment of derived full length amino acid sequence for guava, green pepper, and banana 13-hydroperoxide lyases. The numbering system used is that of guava 13-hydroperoxide lyase. Amino acid residues that are identical in all

three sequences are indicated by pale grey boxes. Similar amino acids are indicated by dark grey boxes. Start sites are indicated by open boxes with solid lines. Deletions/insertions are indicated by black boxes.

Figure 2 shows the complete cDNA sequence and derived amino acid sequence for guava 13-hydroperoxide lyase. The Met-1, Met-6, Met-9, and Met-13 start sites are indicated. Also indicated are peptides that correspond to the HPLC peaks 12, 13, and 15, and the cysteine of the heme binding site at residue 450.

## DETAILED DESCRIPTION OF THE INVENTION

As used in the specification and in the claims, "a" can mean one or more, depending upon the context in which it is used.

The present invention provides a guava fatty acid 13-HPOL and a nucleic acid encoding the lyase. In particular, it provides a guava-derived protein having 13-HPOL function and a nucleic acid encoding such polypeptide or protein. The term "protein" refers to a polymer of amino acids and can include full-length proteins and polypeptides and fragments thereof.

In the present invention, "lyase" means a protein having at least one lyase function. In particular, the term "13-hydroperoxide lyase" and "functional 13-hydroperoxide lyase" means a lyase protein having at least one function exhibited by native 13-hydroperoxide lyase. For example, 13-HPOL function can include the catalytic activity of cleaving a fatty acid 13-hydroperoxide into a C-6 aldehyde and a C-12-ω-oxoacid moiety. Additionally, the protein can have the following characteristics of 13-HPOL: antigenic determinants, binding regions, or the like. A lyase can comprise additional amino acids, such as amino acids linked to the N-terminal end, or amino acids linked to the C-terminal end or amino acids inserted within the lyase sequence, as long as the resulting protein or peptide retains a lyase function.

The 13-HPOL was purified to apparent homogeneity from guava fruit, and the nucleotide sequence for the full-length gene was determined to be 1467 base pairs (SEQ ID NO:10). The translated sequence encodes a total of 488 amino acids (SEQ ID NO: 6), corresponding to a protein with a calculated molecular weight of 54,817 Daltons, a molar extinction coefficient (at 280nm) of 63,590 ± 5% and an isoelectric point of 7.29.

As shown in Figure 1, the derived full length amino acid sequence shows a degree of homology to the 13-HPOL gene that was recently cloned from green pepper (*Capsicum annuum*) (Matsui et al., *FEBS Lett.*, 1996) and banana (*Musa sp.*) (European Patent Application, Publication No. EP 081133 A2). Taking into account deletions and insertions, the alignment reveals that, of the amino acids that overlap with the green pepper sequence (480), 324 are identical and a further 40 are similar (similar amino acids are S, T; D, E; N, Q; R, K; I, L, M, A, V; F, Y, W; whereas G, C, P and H are not considered to have equivalents). This means that the green pepper amino acid sequence and the full length guava amino acid sequence have an identity (homology) of only approximately 67% and a similarity of only approximately 76%. Of the amino acids that overlap with the banana sequence (483), 280 are identical and a further 48 are similar. This means that the banana and guava sequences have an identity (homology) of only approximately 58% and a similarity of only approximately 68%.

There are significant differences between the guava 13-HPOL and the green pepper and banana 13-HPOLs (Figure 1). Comparison of the amino acid sequences for the three proteins shows that both the pepper and banana sequences are shorter than the full length guava sequence. Moreover, the guava gene contains four possible start sites within the first 13 amino acids (methionines 1, 6, 9 and 13), whereas the pepper sequence has only two, corresponding to guava-Met9 and guava-Met13, and the banana sequence has three, one at residue 8 and two corresponding to guava-Met6 and guava-Met9. In addition, the guava sequence contains a unique region at residues 16-22 (T Y P P S L S) (SEQ ID NO: 1), which both the green pepper and banana sequences lack. The unique region can further include residues 16-23 (T Y P P S L S P) (SEQ ID NO: 20); 16-25 (T Y P P S L S P P S) (SEQ ID NO: 21); 16-27 (T Y P P S L S P P S S P)

(SEQ ID NO: 22); 16-28 (T Y P P S L S P P S S P R) (SEQ ID NO: 23); 16-29 (T Y P P S L S P P S S P R P) (SEQ ID NO: 24). In addition, other amino acid and nucleotide sequences encoding 13-HPOL are unique to guava.

Thus, the present invention provides an isolated protein comprising a fatty acid 13-hydroperoxide lyase comprising the amino acid sequence set forth in SEQ ID NO:1, i.e., comprising the amino acid residues found at residues 16-22 of the guava 13-HPOL. The lyase, in addition to having the amino acid residues set forth in SEQ ID NO: 1, can comprise additional amino acid residues so long as the protein retains its lyase function. Examples of such lyases include the fatty acid 13-hydroperoxide lyase isolated from *Psidium guajava*, such as those set forth in SEQ ID NO:2 (guava 13-HPOL Met13), SEQ ID NO:3 (guava 13-HPOL Met9), SEQ ID NO:4 (guava 13-HPOL Met6), and SEQ ID NO:6 (guava 13-HPOL Met1).

It should be noted that active lyase enzyme is obtained upon expression of the guava protein with the sequence including all four of the methionines (guava-Met1), or with the shorter sequences including three (guava-Met6), two (guava-Met9), or only one methionine (guava-Met13). Enzyme activity is three-fold higher when only two (guava-Met9) or one (guava-Met13) methionines are included in the expressed protein.

The present invention additionally provides a fatty acid 13-hydroperoxide lyase comprising at its N-terminus the first eight amino acids of the guava 13-HPOL, i.e., the amino acid sequence set forth in SEQ ID NO:5. For example, the invention provides a protein having the amino acid sequence set forth in SEQ ID NO:6. The term "at its N-terminus" refers to the amino acid residues at the amino terminus of the full length lyase, wherein there may be additional residues attached to the amino terminus of the full length protein. More specifically, the amino acid sequence of the fatty acid 13-hydroperoxide lyase with the amino acid sequence of SEQ ID NO:5 at its N-terminus can be an amino acid sequence present in fatty acid 13-hydroperoxide lyase isolated from *Psidium guajava*.

As will be appreciated by those skilled in the art, the invention also includes those proteins having slight variations in amino acid sequences or other properties. Such variations may arise naturally as allelic variations (*e.g.*, due to genetic polymorphism) or may be produced by human intervention (*e.g.*, by mutagenesis of cloned DNA sequences), such as induced point, deletion, insertion and substitution mutants. When such variations occur, minor changes in amino acid sequence are generally preferred, such as conservative amino acid replacements, small internal deletions or insertions, and additions or deletions at the ends of the molecules. Substitutions may be designed based on, for example, the model of Dayhoff, *et al.*, 1978. These modifications can result in changes in the amino acid sequence, provide silent mutations, modify a restriction site, or provide other specific mutations.

The present invention further provides isolated nucleic acids comprising nucleic acids encoding the proteins of the present invention. For example, the nucleic acid set forth herein as SEQ ID NO: 7 provides a nucleotide sequence for a nucleic acid that encodes the 13-HPOL comprising the amino acid set forth in SEQ ID NO:1, i.e., comprising the amino acid residues found at residues 16-22, 16-23, 16-25, 16-27, 16-28, or 16-29 of the guava 13-HPOL and, more specifically, the guava 13-HPOL Met13 as set forth in SEQ ID NO:2. Other examples of such nucleic acids are the nucleic acids having the nucleotide sequence set forth herein as SEQ ID NO:8, which encodes the guava 13-HPOL Met9 set forth in SEQ ID NO: 3 and SEQ ID NO:9, which encodes the guava 13-HPOL Met6 set forth as SEQ ID NO:4. Yet another example is SEQ ID NO:10, which encodes the protein comprising a fatty acid 13-HPOL comprising at its N-terminus the amino acid sequence set forth in SEQ ID NO:5 and more specifically the guava 13-HPOL Met1 as set forth in SEQ ID NO:6. Additional nucleic acids encoding these proteins can readily be made, utilizing the degeneracy of the genetic code. Additionally, a nucleic acid encoding any selected protein can readily be made, based upon the genetic code, as known in the art. Nucleic acids can be obtained by any of several means known in the art. For example, cDNAs can be isolated from a library using a probe derived from the present nucleic acids or polypeptides, or nucleic acids

can be directly synthesized mechanically. The nucleic acids can be double or single-stranded depending upon the purpose for which it is intended.

The present invention further provides an isolated nucleic acid which specifically hybridizes with the nucleic acid of SEQ ID NO:7 (i.e., the nucleotide sequence encoding guava 13-HPOL Met13 as set forth in SEQ ID NO:2) under stringent conditions of hybridization and which does not hybridize at the stringent conditions to the nucleic acid set forth in SEQ ID NO:11 (i.e., the nucleotide sequence of green pepper 13-HPOL) or SEQ ID NO:12 (i.e., the nucleotide sequence of banana 13-HPOL). Preferably, the isolated nucleic acid has at least 99, 98, 97, 95, 90, 85, 80, 75, or 70% complementarity with the sequence to which it hybridizes. More preferably, the isolated nucleic acid encodes a functional 13-HPOL. The nucleic acid can also be a probe or a primer, for example, to detect or amplify target nucleic acids. Typically, a unique nucleic acid useful as a primer or probe will be at least about 20 to about 25 nucleotides in length, depending upon the specific nucleotide content of the sequence. Additionally, fragments can be, for example, at least about 30, 40, 50, 75, 100, 200, or 500 nucleotides in length. Alternatively, a full length sequence or a sequence that is longer than a full length sequence can be used.

"Stringent conditions" refers to the hybridization conditions used in a hybridization protocol or in the primer/template hybridization in a PCR reaction. In general, these conditions should be a combination of temperatures and salt concentrations for washing chosen so that the denaturation temperature is approximately 5-20°C below the calculated $T_m$ (melting/denaturation temperature) of the hybrid under study. The temperature and salt conditions are readily determined empirically in preliminary experiments in which samples of reference nucleic acid are hybridized to the primer nucleic acid of interest and then amplified under conditions of different stringencies. The stringency conditions are readily tested and the parameters altered are readily apparent to one skilled in the art. For example, $MgCl_2$ concentrations used in PCR buffer can be altered to increase the specificity with which the primer binds to the template, but the concentration range of this compound used in hybridization reactions

is narrow, and therefore, the proper stringency level is easily determined. For example, hybridizations with oligonucleotide probes 18 nucleotides in length can be done at 5-10°C below the estimated $T_m$ in 6X SSPE, then washed at the same temperature in 2X SSPE. The $T_m$ of such an oligonucleotide can be estimated by allowing 2°C for each A or T nucleotide, and 4°C for each G or C. An 18 nucleotide probe of 50% G+C would, therefore, have an approximate $T_m$ of 54°C. Likewise, the starting salt concentration of an 18 nucleotide primer or probe would be about 100-200 mM. Thus, stringent conditions for such an 18 nucleotide primer or probe would be a $T_m$ of about 54°C and a starting salt concentration of about 150 mM and modified accordingly by preliminary experiments. $T_m$ values can also be calculated for a variety of conditions utilizing commercially available computer software (e.g., OLIGO®).

Modifications to the nucleic acids of the invention are also contemplated as long as the essential structure and function of the protein encoded by the nucleic acids is maintained. Likewise, fragments used as primers can have substitutions, so long as enough complementary bases exist for selective amplification, and fragments used as probes can have substitutions, so long as enough complementary bases exist for hybridization with the reference sequence to be distinguished from hybridization with other sequences.

Probes of this invention can be used, for example, to screen genomic or cDNA libraries or to identify complementary sequences by Northern and Southern blotting. Primers of this invention can be used, for example, to transcribe cDNA from RNA and to amplify DNA according to standard amplification protocols, such as PCR, which are well known in the art.

The present invention also provides vectors for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising the nucleic acids of the present invention. More specifically, the vector can be a plasmid. Even more specifically, the vector can

comprise a promoter functionally linked to one of the nucleic acids of the present invention. "Vector" means any carrier containing foreign DNA . "Vectors" include but are not limited to plasmids, viral nucleic acids, viruses, phage nucleic acids, phages, cosmids, and artificial chromosomes. The vector will typically contain appropriate sequences for expression of the 13-HPOL.

The present invention also provides cells containing an exogenous nucleic acid comprising one of the nucleic acids of the present invention. More specifically, the cell can be an *Escherichia coli* or yeast cell.

The present invention also provides a method of cleaving a 13-hydroperoxide of linoleic acid or α-linolenic acid into a $C_6$- aldehyde and a $C_{12}$-oxocarboxylic acid comprising contacting the recombinant protein produced by the vector of the present invention with the 13-hydroperoxide, thereby cleaving the 13-hydroperoxide.

Further provided is a method of preparing n-hexanal, 3-(Z)-hexen-1-al, 2-(E)-hexen-1-al, or their corresponding alcohols from 13-hydroperoxy-octadeca-9,11-dienoic acid (13-HPOD) or 13 hydroperoxy-octadeca-9,11,15-trienoic acid (13-HPOT), comprising contacting the 13-hydroperoxy-octadeca-9,11-dienoic acid or the 13 hydroperoxy-octadeca-9,11,15-trienoic acid with the recombinant protein produced by the vector of the claimed invention, thereby converting the 13-hydroperoxy-octadeca-9,11-dienoic acid into n-hexanal or the 13 hydroperoxy-octadeca-9,11,15-trienoic acid into 3-(Z)-hexen-1-al; and either recovering the n-hexanal or 3-(Z)-hexen-1-al; reducing the n-hexanal into n-hexanol or the 3-(Z)-hexen-1-al into 3-(Z)-hexen-1-ol and recovering the hexanol or 3-(Z)-hexen-1-ol; or isomerizing the 3-(Z)-hexen-1-al under temperature and pH conditions effective to obtain 2-(E)-hexen-1-al and either recovering the formed 2-(E)-hexen-1-al or reducing the 2-(E)-hexen-1-al to 2-(E)-hexen-1-ol and recovering the 2-(E)-hexen-1-ol from the medium.

Also provided is a method of expressing a recombinant protein produced by the transformed cell of the invention, comprising optimizing active lyase function of the recombinant protein by culturing the cells in the absence of isopropyl β-D-thiogalactopyranoside. Active lyase function can be further optimized by culturing the cells in the absence of a heme precursor, including, for example, δ-aminolevulinic acid. Active lyase function can be even further optimized by culturing the cells for greater than 24 hours and, preferably, for 48 hours at approximately 20-30°C, and more preferably at 23°C.

## EXAMPLES OF THE INVENTION

The present invention is more particularly described in the following examples, which are intended as illustrative only since numerous modifications and variations therein will be apparent to those skilled in the art.

EXAMPLE 1

### PURIFICATION OF GUAVA 13-HPOL

Methods

Materials

Unless otherwise stated all chemicals were purchased from Fluka and all HPLC columns were purchased from *Pharmacia.* Pectinex Ultra SP-L was from *Novo Nordisk* Ferment. Immobilon CD membranes were from *Millipore.* Quick-Stain was obtained from *Zoion Research Inc.*

SDS-PAGE

SDS-PAGE with 6.5 or 10% separating gels was done with the buffer system of Laemmli (1970). The gels were stained with Coomassie blue.

Protein determination

Protein concentrations were determining using the dye-binding method of Bradford (1976).

Enzyme assays

13-Hydroperoxides from solubilized linolenic and linoleic acids were produced with soybean 13-HPOL (type V, *Sigma*) according to the method of Vick (1991). During the purification, 13-HPOL activity was measured by following the decrease in absorption at 234 nm, which represents the disruption of the conjugated diene system in the fatty acid hydroperoxide substrate. The assay contained 100 $\mu$M of fatty acid 13-hydroperoxide in 1 ml of 100 mM potassium phosphate buffer at pH 6.0 (Vick, 1991).

Although the above assay was convenient, it did not discriminate between 13-HPOL and other enzymes that consume the 13-hydroperoxides, *i.e.* allene oxide synthase (AOS) (Song and Brash, 1991; Song et al., 1993). Another more specific assay for 13-HPOL was used to confirm that the correct enzyme was purified. This coupled assay used alcohol dehydrogenase and NADH to reduce the aldehydes from the lyase reaction to alcohols. This consumed the NADH and resulted in a decrease in absorption at 340 nm (Vick, 1991).

To verify that the purified enzyme was 13-HPOL, the purified enzyme (50 $\mu$l) was incubated for 30 min in 100 mM potassium phosphate buffer pH 6.0 (5 ml) containing 60 $\mu$M 13-(*S*)-Hydroperoxy-9(*Z*), 11(*E*)-octadecadienoic acid (13-HPOD) or 13-(*S*)-Hydroperoxy-9(*Z*), 11(*E*), 15(*Z*)-octadecatrienoic acid (13-HPOT). Volatiles were extracted twice with 1.5 ml diethyl ether. After concentrating the volume to 400 $\mu$l under a stream of nitrogen, 1 $\mu$l was analyzed by both GC (Supelcowax 10, 15 m, 0.53 mm ID column with a temperature gradient 50 to 120°C at 5° $min^{-1}$) and GC-MS.

Preparations containing 13-HPOL activity were stored frozen at -20°C and neutral pH between each column step.

The non-volatile C-12 fragment was analyzed by radio-HPLC. Accordingly, [$^{14}$C]-13-(S)- hydroperoxylinoleic acid (52 mCi/mmole) was incubated as above with a sample of the purified enzyme for 5 min and the products were analyzed on an Ultrasphere $C_{18}$ column (5$\mu$, 6.4 x 250 mm) using CHCN:$H_2$0:acetic acid (glacial) (60:40:0.01, 1.1 ml/min) as solvent. The results showed that the major polar product had a retention time compatible with that expected of 12-oxo-9(Z)-dodecenoic acid and not that of the characteristic α-ketol derivative formed by plant AOSs.

Purification techniques

Guava fruits were peeled and the pericarp (fleshy) tissue chopped into small pieces. Two volumes of extraction buffer (50 mM sodium phosphate, 1% Triton X-100R, 5 mM sodium ascorbate, pH 7.0) were added to 500 g of chopped pericarp and homogenized for 2 min in a Sorvall mixer at 4°C. The slurry was stirred for 30 min at 4°C. After centrifugation at 16,000 x g for 15 min, 0.02% Pectinex Ultra SP-L solution was added to destroy pectin, and the slurry was stirred for a further 30 min at room temperature to give a preparation referred to as the crude extract (1,600 ml).

Solid $(NH_4)_2SO_4$ was added in small portions to the crude extract at 4°C under stirring until 30% saturation was achieved. After stirring for a further 30 min, the mixture was centrifuged at 20,000 x g for 15 min and the resulting pellet discarded. The supernatant was brought to 60% saturation with more solid $(NH_4)_2SO_4$ added in portions. After stirring for 30 min, the pellet was collected by centrifugation as above. The $(NH_4)_2SO_4$ pellet was dissolved in a minimal volume of extraction buffer (45 ml) and chromatographed (5 runs, maximal loading volume 9 ml) on a Superdex 200 HL 26/60 FPLC gel permeation (GPC) column with 50 mM sodium phosphate, 0.1% Triton X-100R, pH 7.0 as running buffer. The flow used was 2 ml min$^{-1}$, and one fraction was

collected every 2 min. Fractions with 13-HPOL activity (eluting between 75 ml and 85 ml of the run volume) were pooled.

The combined fractions were brought to 30% $(NH_4)_2SO_4$ saturation before loading onto a *Phenyl-Sepharose* HR 26/10 hydrophobic interaction (HIC) column with loading buffer (50 mM sodium phosphate, 1 M $(NH_4)_2SO_4$, pH 7.0). 13-HPOL was eluted with a decreasing salt gradient (100 - 0% over 70 min) with 50 mM sodium phosphate (pH 7.0) containing 0.1% Triton X-100R. The flow used was 8 ml $min^{-1}$ and one fraction was collected each minute. The fractions with 13-HPOL activity (F20- 33) were pooled, concentrated by dialysis against polyethylene glycol 20,000 and then de-salted on a PD-10 column (*Pharmacia*) against the loading buffer (10 mM sodium phosphate, 0.1% Triton X-100R, pH 6.8) for hydroxyapatite chromatography.

The prepared sample was applied to an Econo-Pac HTP column (*Biorad*). 13-HPOL activity was eluted with a gradient from 0-50% of 400 mM sodium phosphate buffer (pH 6.8), containing 0.1% Triton X-100R over 30 min. The flow was 1 ml $min^{-1}$, and one fraction $min^{-1}$ was collected. Fractions with 13-HPOL activity (F13-24) were pooled, concentrated by dialysis against polyethylene glycol 20,000 and then desalted against a loading buffer (75 mM Tris-acetic add, pH 9.3) suitable for isoelectric focusing chromatography (IFC). The prepared sample was applied to a Mono P HR 5/20 column. 13-HPOL activity was eluted with 10% Polybuffer 96 - acetic add, pH 6.0. The flow used was 0.5 ml $min^{-1}$. One fraction was collected every 2 min.

Results

The results of the purification steps are summarized in Table 2.

**Table 2 - Purification of 13-HPOL from Guava Fruit**

| Purification Step | Total Protein (mg) | Total 13-HPOL Activity (nkat) | Recovered Activity (%) | Specific 13-HPOL Activity (nkat $mg^{-1}$) | Purification Factor |
|---|---|---|---|---|---|
| Crude extract | 1,111 | 172,050 | 100.0 | 155 | – |
| 30-60% $(NH_4)_2SO_4$ pellet | 762 | 62,300 | 36.2 | 82 | – |
| GPC | 39 | 32,640 | 19.0 | 837 | 5.4 |
| HIC | 16 | 15,160 | 8.8 | 947 | 6.1 |
| Hydroxylapatite | 1.6 | 6,500 | 3.8 | 4,062 | 26.2 |
| IFC | 0.03 | 317 | 0.2 | 10,566 | 68.2 |

SDS-PAGE analysis of samples from the purification showed that the sample after the chromatofocusing step contained just one, apparently homogenous, band with an apparent molecular weight of 50 kD.

The results from the GPC step indicated that the guava 13-HPOL had a molecular mass of 200 kD. This result, when taken together with the SDS-PAGE analysis of subunit size (50 kD), suggests that the enzyme is a homotetramer. This structure is consistent with data reported for soybean 13-HPOL but inconsistent with data for the enzyme from green pepper fruits and tomato, indicating that the enzymes from green pepper and tomato are trimeric.

The 13-HPOL purified from guava fruit tissue had a broad pH optimum of around 6.0-8.0 and a pI of 6.8 as determined by chromatofocusing.

EXAMPLE 2

TRYPTIC DIGEST AND AMINO ACID SEQUENCE DETERMINATION

Methods

Fractions of purified 13-HPOL were concentrated and then separated on a 6.5% SDS-polyacrylamide gel. Following electrophoresis the proteins were electrotransferred to an ImmobilonCD membrane using a transfer buffer consisting of 10 mM CAPS containing 10% (v/v) methanol pH 11.0. Transfer was achieved in 75 min using a current of 0.8 mA $cm^{-2}$. Proteins were detected by staining using Quick-Stain according to the manufacturer's instructions.

Direct N-Terminal sequencing of the purified 13-HPOL sub-units by *Edman* degradation was not possible as the ends were blocked. The protein band, therefore, was cut out and incubated in 10 $\mu$l of 0.1 M Tris pH 8.2 containing 1 M NaCl , 10% (v/v) acetonitrile, 2 mM $CaCl_2$ and 0.1 $\mu$g trypsin (which cleaves specifically on the carbonyl side of lysine- and arginine-containing peptide linkages) at 37°C for 15 h. After acidification with 1 $\mu$l of 10% TFA the solution was injected directly onto the HPLC system (RP-300 column, *Brownlee*). Chromatography solvents were 0.05% TFA and 2% acetonitrile in water (solvent A) and 0.045% TFA and 80% acetonitrile in water (solvent B). The gradient and flow used were 0 - 5 min 80 $\mu$l $min^{-1}$ at 2% solvent B; 5 - 65 min 50 $\mu$l $min^{-1}$ at 2 - 65% solvent B, and 65 - 70 min 50 $\mu$l $min^{-1}$ at 65 - 100% B. $A_{214}$ was measured in a 200 nl flow cell with a path length of 2 mm. The HPLC-MS interface and post column flow splitting are described in Hess et al., 1993. HPLC-separated peptides were collected manually for sequence analysis and applied to pre-cycled polybrene-treated glass fibre discs. Automated sequencing was done on a model 477A pulsed-liquid phase sequencer (Applied Biosystems, Foster City, California) equipped with a model 120A analyser.

Results

The following amino acid sequence information was obtained for three individual peptides represented by peaks 12, 13 and 15 of the HPLC analysis: Peak Number

12:Asp-Gly-Asn-Ala-Ser-Val-Ile-Phe-Pro-Leu-Gln (SEQ ID NO: 13); Peak Number 13:Asn-Phe-Ala-Met-Asp-Ile-Leu (SEQ ID NO:14); Peak Number 15:Phe-Leu-Phe-Asn-Phe-Leu-Ser (SEQ ID NO:15).

EXAMPLE 3

## DETERMINATION OF THE NUCLEOTIDE AND DERIVED AMINO ACID SEQUENCES OF GUAVA 13-HPOL

Methods

General DNA Manipulation Methods

All media preparation, agarose gel electrophoresis, and general cloning methods were carried out according to standard methods widely known in the art (*Molecular Cloning*, eds. Sambrook, Fritsch, and Maniatis, 1989) unless otherwise stated. QIAprep plasmid kits used for minipreps and QIAquick PCR purification kits were purchased from Quiagen Ltd and used according to the manufacturers' instructions. DNA sequencing was performed using Prism cycle sequencing reagents from Perkin Elmer Ltd and an ABI 373 automatic sequencer. RT- PCR was carried out using an Access RT-PCR System kit purchased from Promega UK according to the manufacturer's instructions. PCR products were cloned using a pGEM-T Vector kit purchased from Promega UK and used according to the supplied instructions.

Degenerate synthetic oligonucleotides were designed and synthesized based on the sense (S) and antisense (A) reverse translations of the three sequences isolated after proteolytic digestion of purified 13-HPOL (see Example 2). The degenerate oligonucleotides were used (1) to determine the arrangement of these three peptides in the primary structure of the enzyme and (2) to generate DNA fragments corresponding to the sequences between the peptides.

Isolation of genomic DNA

Frozen leaf material (5 g) was crushed into a powder in liquid nitrogen in a pre-cooled pestle and mortar. The nitrogen was allowed to evaporate and the powder transferred to a *Dounce* homogenizer containing CTAB buffer (200 mM Tris-Cl pH 8.0, 20 mM EDTA, 1.4 M NaCl, 2% hexadecyltrimethylammonium bromide (CTAB) (w/v), 1% PVP 40,000 (w/v), 28 mM 2-mercaptoethanol v/v, 20 ml). Several strokes of the homogenizer were required to homogenize the powder. The homogenate was transferred to a *Falcon* tube (50 ml) and incubated at 65°C for 90 min. Chloroform: isoamyl alcohol (10 ml, 24:1 v/v) was added and mixed in. The mixture was centrifuged at 3,000 *g* for 60 min which resulted in 3 layers. The upper aqueous layer was transferred to a fresh tube and an equal volume of isopropanol added and gently mixed. Following centrifugation at 3,000 *g* for 20 min, the supernatant was discarded. The pellet was washed with ethanol: 200 mM ammonium acetate (7:3, v/v, 25 ml) and centrifuged as before. The final pellet was resuspended in TE buffer (10 mM Tris-Cl pH 7.5, 1 mM EDTA, 0.5 ml) by heating at 65°C. RNase was added to a final concentration of 600 ng/ml and the solution incubated at 37°C for 1 hr. The yield and purity were determined spectroscopically and by agarose gel electrophoresis.

Isolation of total RNA from guava fruit

Plant material (1 g) was crushed to a fine powder in liquid nitrogen in a pre-cooled pestle and mortar. The nitrogen was allowed to evaporate and the powder transferred to a *Dounce* homogenizer. Lysis buffer (200 mM Borax, 30 mM EGTA, 10 mM DTT, 1% w/v SDS, 1% w/v sodium deoxycholate, 2% PVP 40,000, 0.5% v/v NP-40, 5 ml at 80°C) was added and the powder homogenized. The homogenate was transferred to a Universal flask (30 ml), and proteinase K (125 $\mu$1 @ 20 mg/ml) added. This mixture was incubated at 42°C for 90 min with shaking sufficient to mix the contents without excessive foaming. Aliquots of the mixture (1ml) were transferred to microcentrifuge tubes (1.5 ml) and a solution of KC1 (1 M, 190 $\mu$1) added. After mixing, the tubes were incubated on ice for 1 h and then centrifuged for 10 min.

Aliquots (0.5 ml) were transferred to fresh microcentrifuge tubes and LiCl (4 M, 0.5 ml) added. Following mixing, the tubes were incubated at 4°C overnight. After centrifugation for 10 min, the supernatant was discarded. The pellets were washed with LiCl (2 M, 200 $\mu$l) and centrifuged as before. The supernatant was discarded and the pellets resuspended in TE buffer (200 $\mu$l). The samples were pooled and the RNA (970 µg) quantified by UV spectrophotometry.

Purification of messenger RNA (mRNA) from total RNA

mRNA (18.8 $\mu$g, 1.9% yield) was purified from the total RNA by using an mRNA purification kit *(Pharmacia Biotech)* as described by the manufacturer. The kit uses spun columns of oligo(dT)-cellulose that bind the polyadenylated RNA (mRNA) by affinity interaction.

Construction and screening of a cDNA library from immature guava fruit

Construction of the cDNA library was carried out using the ZAP-cDNA Gigapack II Gold Cloning Kit from Stratagene Ltd. following the provided protocols. Accordingly, a sample of total RNA (831 $\mu$g) was prepared from immature guava fruits (2 g) as described above. The mRNA (8.6 $\mu$g) was isolated (see above) from the majority of this material (670 $\mu$g). The yield (1.2%) is in agreement with those from other eukaryotic sources. Five ng of this mRNA was used to construct a cDNA library in the directional vector λZAP (Stratagene, Cambridge). The original library of 7.3 x 10 clones was amplified to give a stable stock of phage at ~5x 10 plaque forming units/ml (pfu/ml).

Polymerase chain reaction (PCR)

Several sets of conditions were used to perform PCR depending on the template and oligonucleotides used as well as the number of cycles and the temperatures of the various steps. The reaction conditions were the same in all cases with the only

variations being in template concentration, oligonucleotide concentration, number of cycles, temperatures used in each cycle and total volume of the reaction. The conditions used were: 50 mM KCl, 10 mM Tris-Cl pH 9.0 (at 25°C), 0.1 % Triton X- 100 (v/v), 1 mM $MgCl_2$, 200 $\mu$M dNTP's, 25 U/ml *Taq* DNA polymerase.

The following cycle parameters and template and oligonucleotide concentrations for the various PCRs were as follows: genomic DNA with degenerate oligonucleotides (0.5 µg template; 1nmole oligonucleotide; 60 cycles of start at 94°C for 1 min duration, of annealing at 45°C for 1 min, of elongation at 72°C for 1 min); pGEM13-15 with degenerate oligonucleotides (2µl miniprep DNA template; 0.4 nmoles oligonucleotide concentration; 20 cycles of start at 94°C for 1 min, of annealing at 45°C for 0.5 min, of elongation at 72°C for 0.5 min); genomic /pGEM13-15 with Guv13&Guv15a (0.5 µg or 1µl miniprep DNA template; 1 nmole oligonucleotide; 30 cycles of start at 94°C for 1 min, of annealing at 50°C for 1 min, of elongation at 72°C for 1 min). Conditions similar to those used for genomic /pGEM13-15 with Guv13&Guv15a were used for the following except as specifically noted: RACE PCRs (either 1µl of λ DNA or 5 µl of λ phage supernatant with 30 and 60 cycles, respectively, and with 0.01 nmoles of oligonucleotide); screening RACE clones (1 µl miniprep DNA, 0.05 nmoles oligonucleotide, and 10 cycles); nested PCRs (5 µl cleaned PCR as template, 0.1 nmoles oligonucleotide, annealing temperature of 54°C, and 20 cycles); screening PCR (1 µl miniprep DNA, 0.1 or 0.05 nmoles oligonucleotide, annealing temperature of 54°C for 10 cycles); PCRs with the second set of degenerate oligonucleotides (0.5 µg genomic DNA or 5 µl λ phage supernatant, 0.25 oligonucleotide concentration, and 60 cycles). In control reactions, either oligonucleotides or template were omitted, but the corresponding volume of water was added.

Results

PCR was performed using single oligonucleotides or six pairs of oligonucleotides both with and without genomic DNA as template. The pairs of oligonucleotides were

termed 12S&13A, 12S&15A, 13S&15A, 13S&12A, 15S&12A and 15S&13A. For each of the 6 possible arrangements of the peptides in the primary structure of the enzyme, one would expect a different set of PCR products from the reactions. For example, if the arrangement of the peptides was 12-13-15 then the 12S&13A, 12S&15A and 13S&15A PCRs would give products with the 12S&15A product being almost the same size as the other two products added together. After some degree of optimization, products were visible in these reactions and unique products could be observed on a 2% agarose gel of these reaction products.

The agarose gel showed only 2 unique bands, in the 13S&12A and 13S&15A lanes. The 13S&15A product was bigger than the 13S&12A product. This suggested that the orientation of the peptides was 13-12-15. From this, it was expected that a unique product would be observable in the 12S&15A reaction. The product observed was smaller than either of the other two unique products, suggesting that the arrangement of the peptides was 13-12-15, with 12 and 15 being very close together. Calculating the sizes of the PCR products from their mobility on an agarose gel gave the following sizes: 13S&15A, ~160 bp; 13S&12A, ~140 bp; and 12S&15A, ~50 bp. The size of the unique products indicated that the sequence data would not be a major part of the gene sequence.

The 13S&15A unique product was purified from an agarose gel and used in a ligation with pGEM-T (*Promega,* Southampton), a vector designed for the efficient cloning of PCR fragments by the T-tail method (See Promega technical Bulletin TB 150). Two clones were picked and used as template in PCRs with 13S&12A, these clones produced products with the same size as the 13S&12A product observed with the genomic DNA template. This confirmed that the cloned 13S&15A fragments did contain the DNA sequence of peptide 12. The resulting plasmid was named pGEM13-15. Plasmid DNA was sequenced and the amino acid sequence derived (codons 151 to 204 in Figure 2).

This sequence confirmed the following assumptions: (1) the three peptides were from the same peptide chain; (2) the cloned DNA fragment was part of the gene encoding these peptides since the coding sequences for all three were present as part of an open reading frame; and (3) the peptides were very close within the primary structure of the enzyme. Each peptide was cleaved at a lysine residue, as expected from the use of trypsin during the proteolytic digestion of the purified lyase. See Example 2.

A search of the SWISS-PROT and PATTCHX protein databases with the amino acid sequence from this fragment failed to show any protein with a high level of similarity. The degree of dissimilarity between the guava derived sequence and allene oxide synthases suggested that the cloned sequence did not come from a guava allene oxide synthase but from the desired lyase gene. The cloned sequence allowed the design of oligonucleotides that were specific for the lyase gene. The sequences chosen were from within the determined sequence rather than from the ends since the ends were derived from the degenerate oligonucleotides and thus did not necessarily represent the actual guava DNA sequence.

EXAMPLE 4

MOLECULAR CLONING OF THE GENE ENCODING GUAVA 13-HPOL

Results

Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)

First strand cDNA was synthesized from either total or poly(A)+RNA using an oligo-d(T)- adaptor (Table 3, summarizing reaction conditions). The first strand cDNA was then used directly in PCR reactions without purification (Table 4, summarizing reaction conditions). The reaction conditions were the same in all cases except for the amount of template used and DNA polymerase (either AmpliTaq DNA polymerase (*Perkin Elmer*) or Expand High Fidelity (*Boehringer Mannheim*)).

**Table 3 - Reaction Conditions Used for Reverse Transcriptase (1 hour at 30°C)**

| Component | Amount Used |
|---|---|
| Total RNA form immature fruit or mRNA pre-treated at 65°C for 5 min | 5 µg<br>1 µg |
| Oligo-dT adaptor (5'ATG AAT TCG GTA CCC GGG ATC CTT TTT TTT TTT TTT TTT3') | 80 pmoles |
| 5 x first strand buffer | 10 µl |
| DTT | 1 mM |
| dNTP | 1 mM for each |
| RNAsin | 50 units |
| M-MLV-RT | 400 units |
| Water | to 50 µl |

**Table 4 - Reaction Conditions Used for PCR**

| Component | Amount Used |
|---|---|
| guava cDNA (added at 80°C) | 20-100 ng |
| dNTP | 200 µM for each |
| KCl | 50 mM |
| MgCl2 | 3 mM |
| Tris-HCl | 10 mM, pH 8.3 |
| upstream primer | 20 pmoles |
| downstream primer | 20 pmoles |
| Taq polymerase | 1.25 units |
| Water | to 50 µl |
| Reaction cycle parameters | -94°C, 2 min; 1 cycle<br>-57 to 62°C, 1 min; 72°C, 1 min;<br>94°C, 1 min; 30 cycles<br>-72°C, 10 min; 1 cycle |

Synthesis of Lyase Specific Oligonucleotide Primers

Specific oligonucleotide sequences corresponding to sections of the nucleotide sequence of the partial guava clone described in Example 3 were synthesized using methods known in the art (Sambrook et al., 1989).

5' Rapid Amplification of cDNA Ends (5' RACE)

Specific RNA sequences from mRNA (1 $\mu$g) were converted into first strand cDNA using a 5' RACE system (GibcoBRL) as described by the manufacturer.

Cloning and sequencing of the 3' end of the transcript using 3' RACE

The 3' RACE (Rapid Amplification of cDNA Ends) method utilizes a gene-specific upstream primer for PCR, and a downstream primer based on the "adaptor sequence" at the 5' end of the primer used in the reverse transcriptase-catalyzed synthesis of the cDNA. The cDNA was prepared using total guava RNA (5 $\mu$g) and a gene-specific primer.

3'RACE reactions were carried out at an annealing temperature of 57°C. The first round PCR was primed with guava cDNA (1 $\mu$l, corresponding to an original 0.1$\mu$g of total RNA). A gene-specific upstream primer was used with an oligo-adaptor downstream primer. The oligo-dT anneals to the poly-A tail of the mRNAs and the "adaptor" part is another 20 bases of known sequence tagged on the 5' tail of the primer. This sequence was later used as downstream primer for 3' RACE, while the upstream primer was based on the sequence identified in Example 3. No band was obtained when the reaction products were run an agarose gel.

A second round PCR reaction was primed with the first round reaction products (0.1 $\mu$1) and a gene-specific primer as a nested upstream primer (a nested primer is one that corresponds to a sequence within those used in the first round). The downstream

primer was either the oligo-adaptor or one of two gene-specific primers, termed A7205 or A7206, the latter two being complementary to the sequence of the putative 3' UTR (UnTRanslated sequence) obtained as in Example 3.

A third round PCR reaction was primed with the second round product (0.1 $\mu$l, amplified with A7099 and the oligo-adaptor as primer). A third gene-specific nested upstream primer (A7203) and the oligo-adaptor were used as primers.

Cloning and sequencing of the 5' end of the transcript using 5' RACE

Given that the cloning of the 3' end of the transcript required three rounds of PCR and that there was an appreciable amount of contamination of the total RNA preparation with putative genomic DNA, fresh RNA was prepared and then a poly-$A^+$ selection was performed prior to cDNA synthesis for 5' RACE. The hot borate extraction procedure (Wan and Wilkins, 1994) was used to recover total RNA. This proved far superior to the standard method based on guanidinium thiocyanate and phenol-chloroform (Chomczynski and Sacchi, 1987). Poly-$A^+$ selection was carried out as described above.

The cDNA synthesis for 5' RACE was accomplished using a kit (GibcoBRL) as described by the manufacturer. This technique facilitates the isolation and characterization of 5' ends from low-copy messages as it utilizes a gene-specific primer for first strand cDNA synthesis.

First strand cDNA synthesis was primed using the lyase gene-specific antisense oligonucleotide A7204. This permitted cDNA conversion of the mRNA (1 $\mu$g) from immature guava fruits. The first strand cDNA product was purified and then reacted with a terminal deoxynucleotide transferase enzyme (TdT, GibcoBRL) to add homopolymeric dC tails to the 3' ends of the first strand cDNA.

The tailed cDNA was then amplified by PCR using another lyase-specific oligonucleotide and an anchor primer that allowed amplification from the homopolymeric tail. The annealing temperature used in these PCR reactions was 60°C.

Cloning and sequencing of the full-length cDNA

Gene-specific primers were synthesized to correspond to the putative start of the coding sequence (two different methionines were selected) and at the stop codon. A Kozak consensus sequence for translation initiation was included in the upstream primer (Kozak, 1989). In addition to this, the restriction sites *Bam*HI and *Eco*RI were incorporated at the 5' and 3' ends respectively for future sub-cloning work.

The primers for this work were ordered with the DMT (dimethoxy-trityl) protecting groups still in place. They were purified by HPLC (Brash et al., 1996), then deprotected and quantified by UV spectroscopy prior to use in PCR.

The PCR reaction was primed with guava cDNA prepared from guava mRNA (1 $\mu$g) and the lyase-specific primers in the following combinations 1) B6966 with B6967 and 2) C1914 with B6967. The annealing temperature used was 60°C. For these PCR reactions, a special DNA polymerase mixture with proof-reading capabilities was used (Expand High Fidelity, *Boehringer Mannheim*).

Both PCR reactions, i.e., using the two different upstream primers, gave a band of the expected size (1.5 kb). The two different products were subcloned into the pCR2.1 vector (*Invitrogen*) and sequenced.

Results

Cloning and sequencing of the 3' end of the transcript using 3' RACE

The second round PCR reaction gave a unique band with the primer A7205 (250 bp) and also a unique band (220 bp) with A7206. The difference in size of these two PCR products (30 bp) matches the expected distance between the two downstream primers A7205 and A7206. Furthermore, the sizes of 220 and 250 bp were exactly what was expected by direct cloning of the DNA fragment derived in Example 3.

The 220 or 250 bp product was, however, too short to encode the full length of the remaining 3' coding sequence and 3' UTR of the lyase. The expected size of the correct PCR product was at least 950 bp, and it could have significantly longer depending on the length of the 3' UTR. This finding was interpreted as indicating that the original clone was derived from a fragment of genomic DNA that had been cloned into the cDNA library. The coding sequence obtained corresponded to an exon, and this led into an intron (non-coding sequence, originally suspected to be 3'UTR) immediately after the coding sequence for Asn-Ile-Gly.

During the same series of second round PCR reactions, the reaction using the oligo-adaptor downstream primer amplified 2 products (450 and 1,100 bp). The larger product is compatible in size to the expected product from the 3' end of the lyase cDNA.

During the third round PCR reaction, a 1,000 bp product corresponding to the expected size was obtained. This PCR product differed in size from the second round product by 100 bp, which corresponds well to the different positions of the nested upstream primers in these two PCR reactions. This 1,000 bp product was sub-cloned into the pCR2.1 vector (Invitrogen) and sequenced.

The sequence showed that the PCR product contained the sequence identified in Example 3 together with the remainder of the 3' coding sequence plus 186 bp of 3' UTR.

Cloning and sequencing of the 5' end of the transcript using 5' RACE

Primer C1589 gave a unique band of the expected size (500 bp) as did primer C1588 (700 bp). The 700 bp PCR product was sub-cloned into the pCR2.1 vector (Invitrogen) and sequenced.

Cloning and sequencing of the full-length cDNA

The complete sequence of the product of the PCR reactions is shown in Figure 2. The translated sequence encodes a total of 488 amino acids corresponding to a protein with a calculated molecular weight of 54,817 Daltons, a molar extinction coefficient (at 280nm) of 63,590 ± 5%, and an isoelectric point of 7.29.

EXAMPLE 5

EXPRESSION OF THE GENE ENCODING GUAVA 13-HPOL

Bacterial transformation

The full-length cDNA clone of 13-HPOL (See Figure 2) was inserted into the *Escherischia coli* expression plasmid pET30b (*Novagen).* The pET30b system contains a sequence that "tags" the expressed protein with a number of histidine residues. This provides a means of purifying the protein by affinity chromatography using a nickel ligand.

Accordingly, the pET30b plasmids and the pCR2.1 clone containing the 13-HPOL cDNA were linearized with 2 different restriction enzymes (BamHI and HindIII) and then ligated together. The pET30b:13-HPOL constructs were used to transform *E. coli* strain BL21 (Novagen).

Expression of 13-HPOL in transformed *E. coli* cells

The transformed BL21 cells were cultured overnight at 37°C and 280 rpm in LB medium (3 ml, prepared by dissolving tryptone (10 g), yeast extract (5 g), and NaCl (10 g) in 1 liter of water, adjusting the pH to 7.0 and autoclaving). The antibiotic kanamycin (30 mg) was added aseptically after autoclaving. A portion of the resulting culture (0.2 ml) was then transferred to Terrific Broth (TB, 10 ml, prepared by dissolving bacto-tryptone (12 g), bacto-yeast extract (12 g), and glycerol (4 ml) in deionized water (900 ml), autoclaving and then adding a sterile solution (100 ml) containing 50μg/ml kanamycin, 0. 17 M $KH_2PO_4$, and 0.72 M $K_2HP0_4$) and allowed to grow until the optical density at 260 nm ($OD^{260}$) reached 0.6. This culture was used to inoculate 50 ml of TB containing 50 μg/ml of kanamycin, which was then placed at 28°C and 200 rpm and a heme precursor, δ- aminolevulimic acid (1 mM), was added followed by the inducer IPTG (0.4 mM) one hour later. The induced cultures were left for a further period of time (4 or 16 hours) and the cells harvested by centrifugation (5,000 rpm for 7 min at 4°C). The precipitated cells were washed by resuspending them in Tris-HC1 buffer (50 mM, pH 7.9) followed by recentrifugation as before.

The resulting pellet of cells was resuspended in Tris-acetate buffer (0.1 M, pH 7.6) containing sucrose (0.5 M), EDTA (0.5 mM) and lysozyme (1 mg/ml). After 30 min on ice, the mixture was centrifuged as before to obtain a pellet of spheroplastes. These were resuspended in potassium phosphate buffer (0.1 M, pH 7.6) containing magnesium acetate (6 mM), glycerol (20% v/v) and DTT (0.1 mM) and the mixture left for 10 min at -80°C. Following this, a protease inhibitor was added (PMSF, 1 mM) and the cells sonicated (2 x 30 seconds). 13-HPOL activity was readily detected in this sonicate using the methods described herein.

SDS-PAGE analysis of 13-HPOL expression

Proteins from the transformed and induced cells were compared by SDS-PAGE with those from control cultures. The results from the analysis of the pET30b: 13-HPOL

constructs in *E. coli* strain BL21 clearly showed that a huge amount of protein with the expected molecular weight (54 kD) had been made.

EXAMPLE 6

EXPRESSION OF VARIANTS OF THE GENE ENCODING GUAVA 13-HPOL UNDER VARIED CONDITIONS

Four different cDNA clones of 13-HPOL (13-HPOL-Met1, -Met6, -Met9, -Met13) were inserted into the *E. coli* expression plasmid pET30b. The pET30b:13-HPOL were used to transform *E. coli* strain BL21 (Novagen) under various conditions.

Methods

Bacterial strain and plasmid

The bacterial host strain BL21(DE3) : (F- ompT hsdSB (RB-mB-) gal dcm (DE3)) and pET30b plasmids were obtained from Novagen.

Constructs

Four expression plasmids (pET30b : 13HPOL-Met1, -Met6, -Met9 and -Met13) were made according to procedures well known in the art (e.g., Sambrook et al., 1989) . Construct pET30b:13HPOL-Met1was made as follows: cDNA encoding the 13-HPOL-Met1 in pCR2.1 was cut with BamHI and HindIII and subcloned into the expression vector plasmid pET30b (digested also with BamHI and HindIII). The 13-HPOL-Met1 construct was used to transform *E. coli* strain XLI-Blue by heat shock. Colonies obtained after transformation were grown in 2 ml of LB medium containing 30 µg/ml of kanamycin at 37°C overnight and plasmid DNAs were purified using a Qiagen Plasmid Kit. The plasmid DNA was cut with BamHI and HindIII to screen for the correct plasmid DNA, pET30:13-HPOL-Met1. Then, the plasmid DNA was used for transformation of *E. coli* strain BL21(DE3) to express the 13-HPOL.

Constructs pET30b:13-HPOL-Met6, -Met9, -Met13 were made using the construct pET30b:13-HPOL-Met1. PCR products of approximately 680 to 700bp of Met6, Met9, and Met13 with BamHI and MscI cleavage sites were each subcloned into pCR2.1 and subsequently digested with BamHI MscI. The pET30 digestion product of pET30:13-HPOL-Met1 and the PCR digestion product of the preceding step were purified and ligated to form Constructs pET30b:13-HPOL-Met6, -Met9, -Met13.

A PCR reaction was carried out under the following conditions: 20-100ng of cDNA, 1µl dNTP 10mM, 5µl PCR buffer (10x) with 15mM $MgCl_2$, 5 µl of 4µM primer downstream, 0.75 Expand ™ High Fidelity, Boehringer Mannheim, and water up to 50 µl. The PCR buffer (10x; Expand™ High Fidelity) consisted of 20mM Tris-HCl (pH7.5), 100mM KCl, 1mM DTT (dithiothreitol), 0.1mM EDTA, 0.5% (v/v) Tween 20, 0.5% (v/v) Nonidet P40, and 50% (v/v) glycerol. The reaction was primed with the cDNA encoding the 13-HPOL-Met1 in pCR2.1 and using as primer either (1) Guava-up-Met6 ($^{5'}$GCG GAT CCG GCC ATG AGC AAC ATG TCG$^{3'}$) (SEQ ID NO:16) and Guava-down ($^{5'}$AAT GTT GAT GGT GGG GAG GAG$^{3'}$) (SEQ ID NO:17), (2) Guava-up-Met9 ($^{5'}$GCG GAT CCG GCC ATG TCG CCG GCC AT$^{3'}$) (SEQ ID NO:18) and Guava-down, or (3)Guava-up-Met13 ($^{5'}$GCG GAT CCG GCC ATG TCG TCC ACC TAC$^{3'}$) (SEQ ID NO:19).

In each PCR reaction, a unique band was amplified which corresponded to the first 680-700 bp of the 13-HPOL starting from methionine in position 6, 9, or 13. After purification (QIAEX II gel extraction kit), each DNA fragment was subcloned into the vector pCR2.1 and sequenced.

Preparation of bacterial cultures

The bacterial cultures were prepared according to the method of Hoffman et al. (1995). Specifically, a single bacterial colony from a complex agar plate containing 30 µg/ml of kanamycin was grown in 1 ml of LB medium containing 50 µg/ml of

kanamycin for 3 hours at 37°C. A small aliquot (200 µl) of this culture was then used to inoculate 10 ml of LB or TB containing 50 µg/ml of kanamycin and the culture was again grown at 37°C. After 3 hours, this culture was used to inoculate 50 ml of LB or TB medium containing 30 µg/ml of kanamycin and with or without 1 mM of δ-aminolevulinic acid (δ-ALA).

The culture was grown at 15°C, room temperature (23°C) or 28°C. After 1 hour, the inducer isopropyl-β-D-thiogalactopyranoside (IPTG, 0.4 mM) was added or not. The culture was grown 4 hours, 24 hours or 48 hours at 15°C, room temperature (23°C) or 28°C.

The bacterial cells were centrifuged at 4°C for 10 min (5,000 rpm). The precipitated cells were washed by resuspension in 10 ml of Tris-HCl buffer 50mM pH7.9 and were centrifuged as before. Sonicates were prepared as described in Example 5.

The activity of the 13-HPOL expressed in the BL21 cells was measured using the spectrophotometric assay and by HPLC as described above. The sample was diluted 10-fold, and 5 or 10 µl aliquots were assayed using 4-5 mg of 13(S)-hydroperoxylinolenic acid in 0.5 ml of potassium phosphate pH 7.4. The decrease in absorbance at 235 nm was immediately recorded. The activity was verified using GC-MS of the volatile C-6 product as described above. No activity could be detected in the negative control, which consisted of the sonicated protein preparation obtained from the BL21 cells transformed with pET30 only.

Results

The results are shown in Table 5. At 15°C, no lyase activity was detectable. The activity at 23°C was higher than at 28°C. Activities were highest after 48 hours of

culture. The best activities were obtained without δ-ALA and without IPTG. Thus, in the system used here, with the pET30 plasmid and its T7 RNA polymerase promoter, and with the cells grown in a rich medium (TB), addition of heme precursor or IPTG inducer does not help with expression of active lyase.

**Table 5 - Activity of 13-HPOL-Met1 after expression in *E. coli* cells in different conditions of culture**

| | Activity (OD. min$^{-1}$) | | | | | |
|---|---|---|---|---|---|---|
| | 4 hours | | 24 hours | | 48 hours | |
| | 23°C | 28°C | 23°C | 28°C | 23°C | 28°C |
| +δ-ALA -IPTG | 0.55 | 0.32 | 1.9 | 1.01 | 2.35 | 1.13 |
| -δ-ALA -IPTG | 0.7 | 0.45 | 2.34 | 1.44 | 3.01 | 2.6 |
| +δ-ALA +IPTG | - | 0.27 | - | 0.4 | - | - |
| -δ-ALA +IPTG | - | 0.14 | - | 0.24 | - | - |

Proteins from the transformed and induced cells were compared by SDS-PAGE. A very high amount of protein (accounting for more that half of the total cellular protein) with the expected molecular weight was expressed under all the different conditions of culture. Relatively lower amounts of protein, but the highest lyase activity, were obtained when the cells were grown without δ-ALA and IPTG. This may reflect the fact that there is a very high basal induction in this system. Including a further stimulus for protein expression resulted in even lower recovery of correctly folded protein with catalytic activity. In fact, when the cells are examined under the microscope, many inclusion bodies were seen. The number of inclusion bodies was

highest in cells grown with IPTG, in agreement with the concept that the bacteria cannot handle this level of expressed protein.

Substrate specificity of the expressed pET30:13-HPOL-Met1 was examined by incubating an aliquot of sonicated preparation with the following substrates: 13(*S*)-hydroperoxylinoleic acid; 9(*S*)-hydroperoxylinoleic acid; 13(*S*)-hydroperoxylinolenic acid; 15(*S*)-HPETE, which is the 15-hydroperoxide of arachidonic acid. The results showed that the metabolism of the 9(*S*)-hydroperoxylinoleic acid and 15(*S*)-HPETE by the 13-HPOL-Met1 is low in comparison to the rate of reaction with 13(*S*)-hydroperoxylinolenic acid. In addition, the 13-HPOL-Met1 is at least 10-times more active with 13-(*S*)-hydroperoxylinolenic acid as the substrate compared to 13-hydroperoxylinoleic acid.

The expressions of pET30:13-HPOL-Met1,-Met6,-Met9 and -Met13 were also compared. To compare the activities of the four pET30:13-HPOL constructs, the plasmids were expressed under identical conditions. Each was assessed for lyase activity (UV assay) and the level of protein expression (SDS-PAGE). The results from two independent experiments showed that these 4 different enzyme constructs were expressed at similar levels. All four constructs gave active lyase, although with a three-fold range of activities. The Met9 and Met13 gave the highest activities, 0.90 and 0.92 $OD.min^{-1}$ , respectively. Met6 activity was 0.60 $OD.min^{-1}$. Met1 activity was lowest at 0.30 $OD.min^{-1}$.

EXAMPLE 7

## CLEAVAGE OF 13-HPOD TO HEXANAL USING RECOMBINANT GUAVA 13-HPOL

### Methods

A solution of 13-HPOD (55 g/l) was made as described in U.S. Patent No. 5,464,761 and diluted 10 fold with buffer (0.1M potassium phosphate, pH 8.5). Three different quantities of recombinant 13-HPOL were added to the diluted 13-HPOD solution in order to analyze the amount of 13-HPOD cleavage by the recombinant protein. Thus, either 0, 10, or 25 µl of the 13-HPOL containing bacterial lysate (see Example 5) were added to 2 ml of the diluted 13-HPOD solution, which contained 11 mg of 13-HPOD. In samples without bacterial lysate, the lysate was replaced with 10 µl of distilled water. The samples were stirred for 30 min at room temperature (20°C). Each sample was then extracted once with 2 ml of diethylether containing 137 mg/1 n-hexanol as an internal standard. Subsequently, 1 µl of the organic extract was injected onto a 15 M SPwax gas chromatography column using the following temperature program: 50°C for 2 min and ramp to160°C at 5°C each minute. The amount of hexanal formed was calculated by comparison of the hexanal peak area to that of the internal standard.

### Results

The amount of hexanal in the control samples, which contained no 13-HPOL containing bacterial lysate, was 61 mg/l. The hexanal in the control samples was formed by the soy flour used in the preparation of the 13-HPOD substrate. Thus, 61 mg/l was subtracted from the total amount of hexanal in the samples to determine the amount of hexanal formed by the recombinant 13-HPOL in the bacterial lysate. The amount of hexanal formed by the recombinant 13-HPOL in 10 µl of bacterial lysate was 140µg. By extrapolation, one liter of lysate would produce 14 grams of hexanal.

Throughout this application, various publications are referenced. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

Although the present process has been described with reference to specific details of certain embodiments thereof, it is not intended that such details should be regarded as limitations upon the scope of the invention except as and to the extent that they are included in the accompanying claims. Thus, the preceding examples are intended to illustrate, but not limit, the invention. While they are typical of those that might be used, other procedures known to those skilled in the art may be alternatively employed.

## REFERENCES


1. Bradford M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72:248-254.

2. Brash, A.R., Boeglin, W. E. Chang, M. S. and Shieh, B. H. (1996). Purification and Molecular cloning of an 8*R*-Lipoxygenase from the Coral *Plexaura homomalla* Reveal the Related Primary Structures of R- and S-Lipoxygenases. *J. Biol. Chem.*, 271:34:20949-20957.

3. Chomczynski, P. and Sacchi, N. (1987). Single Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction. *Anal. Biochem.* 162, 156-159.

4. Dayhoff, et al., *Atlas of Protein Sequence and Structure*. Nat'l. Biomed. Res. Found., Washington, D.C. (1978)

5. Fauconnier, M.L., Perez, A.G., Sanz, C., Marlier, M. (1997). Purification and Characterization of Tomato Leaf (*Lycopersicon esculentum* Mill.) Hydroperoxide Lyase. *J. Agric. Food Chem.* 45:4232.

6. Hatanaka A. (1993). The biogeneration of green odour by green leaves. *Phytochemistry* 34:1201-1218.

7. Hatanaka A., Kajiwara, T. and Sekija, J. (1987). Biosynthetic pathway for C6--aldehydes formation from linolenic acid in green leaves. *Chemistry and Physics of Lipids* 44:431-361.

8. Hess D., Covey T.C., Winz R., Brownsey R.W. and Aebersold R. (1993). Analytical and micro preparative peptide mapping by high performance liquid chromatography/electrospray mass spectrometry of proteins purified by gel electrophoresis. *Protein Science* 2:1342-1351.

9. Hoffman, et al., (1995) *Protein Expression and Purification* 6: 646-654.

10. Kozak, M. (1989). The Scanning Model for Translation: An update. *J. Cell Biol.* 108:229-241.

11. Laemmli U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685.

12. Matsui K., Shibata Y., Kajiwara, T. and Hatanaka A. (1989). Separation of 13 - and 9-hydroperoxide lyase activities in cotyledons of cucumber seedlings. Z. *Naturforsch.* 44c:883-885.

13. Matsui K, Toyota H., Kajiwara T., Kakuno T. and Hatanaka A. (1991). Fatty acid hydroperoxide cleaving enzyme, hydroperoxide lyase, from tea leaves. *Phytochemistry* 30:2109-2113.

14. Matsui, K., Shibutani, M., Hase, T., and Kajiwara, T. (1996). Bell Pepper Fruit Fatty Acid Hydroperoxide Lyase is a Cytochrome P-450 (CYP74B). *FEBS Lett.* 394:21-24.

15. Olias J.M., Rios J.J., Valle M., Zamora R., Sanz L.C. and Axelrod B. (1990). Fatty add hydroperoxide lyase in germinating soybean seedlings. *J. Agric. Food Chem.* 38:624-630.

16. Sambrook, J., Fritsch, E.F. and Maniatias, T. Eds. (1989) Molecular Cloning. A Standard Laboratory Manual. 2nd Edition. Cold Spring Harbour Laboratory Press.

17. Schreier P. and Lorenz G. (1982). Separation, partial purification and characterization of a fatty acid hydroperoxide cleaving enzyme from apple and tomato fruits. *Z. Naturforsch.* 37c:165-173.

18. Shibata Y., Matsui K, Kajiwara T. and Hatanaka, A. (1995). Purification and properties of fatty acid hydroperoxide lyase from green bell pepper fruits. *Plant Cell Physiology* 36:147-156.

19. Song W.-C. and Brash A.R. (1991). Purification of an allene oxide synthase and identification of the enzyme as a cytochrome P 450. *Science* 253:781-784.

20. Song W.-C., Funk C.D. and Brash A.R. (1993). Molecular cloning of an allene oxide synthase: A cytochrome P-450 specialized for the metabolism of fatty acid hydroperoxides. *Proc. Natl. Acad. Sci. USA* 90:8519-8523.

21. Tressl, R. and Drawert, F. (1973). Biogenesis of banana volatiles. *J. Agric. Food Chem.* 21:560-565.

22. Vick B.A. (1991). A spectrophotometric assay for hydroperoxide lyase. *Lipids* 26:315-320.

23. Vick B.A. and Zimmerman D.C. (1976). Lipoxygenase and hydroperoxide lyase in germinating watermelon seedlings. *Plant Physiol.* 57:780-788.

24. Wan, C.Y., and Wilkins, T.A. (1994). A Modified Hot Borate Method Significantly Enhances the Yield of High-Quality RNA from Cotton *(Gossypium hirsutum L.). Anal. Biochem.* 223:7-12.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Whitehead, Ian M.
Slusarenko, Alan
Wäspi, Urs
Gaskins, Duncan
Brash, Alan
Tijet, Nathalie

(ii) TITLE OF THE INVENTION: Guava (Psidium guajava) 13-Hydroperoxide Lyase and Uses Thereof

(iii) NUMBER OF SEQUENCES: 24

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: NEEDLE & ROSENBERG, P.C.
(B) STREET: 127 PEACHTREE STREET, NE, SUITE 1200
(C) CITY: ATLANTA
(D) STATE: GA
(E) COUNTRY: USA
(F) ZIP: 30303-1811

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Perryman, David G
(B) REGISTRATION NUMBER: 33,438
(C) REFERENCE/DOCKET NUMBER: 06027.0001

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 404 688 0770
(B) TELEFAX: 404 688 9880
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
Thr Tyr Pro Pro Ser Leu Ser
 1               5
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 476 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ser Ser Thr Tyr Pro Pro Ser Leu Ser Pro Pro Ser Ser Pro Arg
 1               5                  10                  15
Pro Thr Thr Leu Pro Val Arg Thr Ile Pro Gly Ser Tyr Gly Trp Pro
            20                  25                  30
Leu Leu Gly Pro Ile Ser Asp Arg Leu Asp Tyr Phe Trp Phe Gln Gly
        35                  40                  45
Pro Glu Thr Phe Phe Arg Lys Arg Ile Glu Lys Tyr Lys Ser Thr Val
    50                  55                  60
Phe Arg Ala Asn Val Pro Pro Cys Phe Pro Phe Phe Ser Asn Val Asn
65                  70                  75                  80
Pro Asn Val Val Val Val Leu Asp Cys Glu Ser Phe Ala His Leu Phe
                85                  90                  95
Asp Met Glu Ile Val Glu Lys Ser Asn Val Leu Val Gly Asp Phe Met
            100                 105                 110
Pro Ser Val Lys Tyr Thr Gly Asn Ile Arg Val Cys Ala Tyr Leu Asp
        115                 120                 125
Thr Ser Glu Pro Gln His Ala Gln Val Lys Asn Phe Ala Met Asp Ile
    130                 135                 140
Leu Lys Arg Ser Ser Lys Val Trp Glu Ser Glu Val Ile Ser Asn Leu
145                 150                 155                 160
Asp Thr Met Trp Asp Thr Ile Glu Ser Ser Leu Ala Lys Asp Gly Asn
                165                 170                 175
Ala Ser Val Ile Phe Pro Leu Gln Lys Phe Leu Phe Asn Phe Leu Ser
            180                 185                 190
Lys Ser Ile Ile Gly Ala Asp Pro Ala Ala Ser Pro Gln Val Ala Lys
        195                 200                 205
Ser Gly Tyr Ala Met Leu Asp Arg Trp Leu Ala Leu Gln Leu Leu Pro
    210                 215                 220
Thr Ile Asn Ile Gly Val Leu Gln Pro Leu Val Glu Ile Phe Leu His
225                 230                 235                 240
Ser Trp Ala Tyr Pro Phe Ala Leu Val Ser Gly Asp Tyr Asn Lys Leu
                245                 250                 255
Tyr Gln Phe Ile Glu Lys Glu Gly Arg Glu Ala Val Glu Arg Ala Lys
            260                 265                 270
Ala Glu Phe Gly Leu Thr His Gln Glu Ala Ile His Asn Leu Leu Phe
        275                 280                 285
Ile Leu Gly Phe Asn Ala Phe Gly Gly Phe Ser Ile Phe Leu Pro Thr
    290                 295                 300
Leu Leu Ser Asn Ile Leu Ser Asp Thr Thr Gly Leu Gln Asp Arg Leu
305                 310                 315                 320
Arg Lys Glu Val Arg Ala Lys Gly Gly Pro Ala Leu Ser Phe Ala Ser
                325                 330                 335
```

```
Val Lys Glu Met Glu Leu Val Lys Ser Val Val Tyr Glu Thr Leu Arg
            340                 345                 350
Leu Asn Pro Pro Val Pro Phe Gln Tyr Ala Arg Ala Arg Lys Asp Phe
        355                 360                 365
Gln Leu Lys Ser His Asp Ser Val Phe Asp Val Lys Lys Gly Glu Leu
    370                 375                 380
Leu Cys Gly Tyr Gln Lys Val Val Met Thr Asp Pro Lys Val Phe Asp
385                 390                 395                 400
Glu Pro Glu Ser Phe Asn Ser Asp Arg Phe Val Gln Asn Ser Glu Leu
                405                 410                 415
Leu Asp Tyr Leu Tyr Trp Ser Asn Gly Pro Gln Thr Gly Thr Pro Thr
            420                 425                 430
Glu Ser Asn Lys Gln Cys Ala Ala Lys Asp Tyr Val Thr Leu Thr Ala
        435                 440                 445
Cys Leu Phe Val Ala Tyr Met Phe Arg Arg Tyr Asn Ser Val Thr Gly
    450                 455                 460
Ser Ser Ser Ser Ile Thr Ala Val Glu Lys Ala Asn
465                 470                 475
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 480 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Met Ser Pro Ala Met Ser Ser Thr Tyr Pro Pro Ser Leu Ser Pro Pro
 1               5                  10                  15
Ser Ser Pro Arg Pro Thr Thr Leu Pro Val Arg Thr Ile Pro Gly Ser
            20                  25                  30
Tyr Gly Trp Pro Leu Leu Gly Pro Ile Ser Asp Arg Leu Asp Tyr Phe
        35                  40                  45
Trp Phe Gln Gly Pro Glu Thr Phe Phe Arg Lys Arg Ile Glu Lys Tyr
    50                  55                  60
Lys Ser Thr Val Phe Arg Ala Asn Val Pro Pro Cys Phe Pro Phe Phe
65                  70                  75                  80
Ser Asn Val Asn Pro Asn Val Val Val Val Leu Asp Cys Glu Ser Phe
                85                  90                  95
Ala His Leu Phe Asp Met Glu Ile Val Glu Lys Ser Asn Val Leu Val
            100                 105                 110
Gly Asp Phe Met Pro Ser Val Lys Tyr Thr Gly Asn Ile Arg Val Cys
        115                 120                 125
Ala Tyr Leu Asp Thr Ser Glu Pro Gln His Ala Gln Val Lys Asn Phe
    130                 135                 140
Ala Met Asp Ile Leu Lys Arg Ser Ser Lys Val Trp Glu Ser Glu Val
145                 150                 155                 160
Ile Ser Asn Leu Asp Thr Met Trp Asp Thr Ile Glu Ser Ser Leu Ala
                165                 170                 175
Lys Asp Gly Asn Ala Ser Val Ile Phe Pro Leu Gln Lys Phe Leu Phe
            180                 185                 190
Asn Phe Leu Ser Lys Ser Ile Ile Gly Ala Asp Pro Ala Ala Ser Pro
        195                 200                 205
Gln Val Ala Lys Ser Gly Tyr Ala Met Leu Asp Arg Trp Leu Ala Leu
    210                 215                 220
Gln Leu Leu Pro Thr Ile Asn Ile Gly Val Leu Gln Pro Leu Val Glu
225                 230                 235                 240
```

```
Ile Phe Leu His Ser Trp Ala Tyr Pro Phe Ala Leu Val Ser Gly Asp
                245                 250                 255
Tyr Asn Lys Leu Tyr Gln Phe Ile Glu Lys Glu Gly Arg Glu Ala Val
            260                 265                 270
Glu Arg Ala Lys Ala Glu Phe Gly Leu Thr His Gln Glu Ala Ile His
        275                 280                 285
Asn Leu Leu Phe Ile Leu Gly Phe Asn Ala Phe Gly Gly Phe Ser Ile
    290                 295                 300
Phe Leu Pro Thr Leu Leu Ser Asn Ile Leu Ser Asp Thr Thr Gly Leu
305                 310                 315                 320
Gln Asp Arg Leu Arg Lys Glu Val Arg Ala Lys Gly Gly Pro Ala Leu
                325                 330                 335
Ser Phe Ala Ser Val Lys Glu Met Glu Leu Val Lys Ser Val Val Tyr
            340                 345                 350
Glu Thr Leu Arg Leu Asn Pro Pro Val Pro Phe Gln Tyr Ala Arg Ala
        355                 360                 365
Arg Lys Asp Phe Gln Leu Lys Ser His Asp Ser Val Phe Asp Val Lys
    370                 375                 380
Lys Gly Glu Leu Leu Cys Gly Tyr Gln Lys Val Val Met Thr Asp Pro
385                 390                 395                 400
Lys Val Phe Asp Glu Pro Glu Ser Phe Asn Ser Asp Arg Phe Val Gln
                405                 410                 415
Asn Ser Glu Leu Leu Asp Tyr Leu Tyr Trp Ser Asn Gly Pro Gln Thr
            420                 425                 430
Gly Thr Pro Thr Glu Ser Asn Lys Gln Cys Ala Ala Lys Asp Tyr Val
        435                 440                 445
Thr Leu Thr Ala Cys Leu Phe Val Ala Tyr Met Phe Arg Arg Tyr Asn
    450                 455                 460
Ser Val Thr Gly Ser Ser Ser Ser Ile Thr Ala Val Glu Lys Ala Asn
465                 470                 475                 480
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 483 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Ser Asn Met Ser Pro Ala Met Ser Ser Thr Tyr Pro Pro Ser Leu
 1                   5                  10                  15
Ser Pro Pro Ser Ser Pro Arg Pro Thr Thr Leu Pro Val Arg Thr Ile
            20                  25                  30
Pro Gly Ser Tyr Gly Trp Pro Leu Leu Gly Pro Ile Ser Asp Arg Leu
        35                  40                  45
Asp Tyr Phe Trp Phe Gln Gly Pro Glu Thr Phe Phe Arg Lys Arg Ile
    50                  55                  60
Glu Lys Tyr Lys Ser Thr Val Phe Arg Ala Asn Val Pro Pro Cys Phe
65                  70                  75                  80
Pro Phe Phe Ser Asn Val Asn Pro Asn Val Val Val Val Leu Asp Cys
                85                  90                  95
Glu Ser Phe Ala His Leu Phe Asp Met Glu Ile Val Glu Lys Ser Asn
            100                 105                 110
Val Leu Val Gly Asp Phe Met Pro Ser Val Lys Tyr Thr Gly Asn Ile
        115                 120                 125
Arg Val Cys Ala Tyr Leu Asp Thr Ser Glu Pro Gln His Ala Gln Val
    130                 135                 140
```

```
Lys Asn Phe Ala Met Asp Ile Leu Lys Arg Ser Ser Lys Val Trp Glu
145                 150                 155                 160
Ser Glu Val Ile Ser Asn Leu Asp Thr Met Trp Asp Thr Ile Glu Ser
                165                 170                 175
Ser Leu Ala Lys Asp Gly Asn Ala Ser Val Ile Phe Pro Leu Gln Lys
            180                 185                 190
Phe Leu Phe Asn Phe Leu Ser Lys Ser Ile Ile Gly Ala Asp Pro Ala
        195                 200                 205
Ala Ser Pro Gln Val Ala Lys Ser Gly Tyr Ala Met Leu Asp Arg Trp
    210                 215                 220
Leu Ala Leu Gln Leu Leu Pro Thr Ile Asn Ile Gly Val Leu Gln Pro
225                 230                 235                 240
Leu Val Glu Ile Phe Leu His Ser Trp Ala Tyr Pro Phe Ala Leu Val
                245                 250                 255
Ser Gly Asp Tyr Asn Lys Leu Tyr Gln Phe Ile Glu Lys Glu Gly Arg
            260                 265                 270
Glu Ala Val Glu Arg Ala Lys Ala Glu Phe Gly Leu Thr His Gln Glu
        275                 280                 285
Ala Ile His Asn Leu Leu Phe Ile Leu Gly Phe Asn Ala Phe Gly Gly
    290                 295                 300
Phe Ser Ile Phe Leu Pro Thr Leu Leu Ser Asn Ile Leu Ser Asp Thr
305                 310                 315                 320
Thr Gly Leu Gln Asp Arg Leu Arg Lys Glu Val Arg Ala Lys Gly Gly
                325                 330                 335
Pro Ala Leu Ser Phe Ala Ser Val Lys Glu Met Glu Leu Val Lys Ser
            340                 345                 350
Val Val Tyr Glu Thr Leu Arg Leu Asn Pro Pro Val Pro Phe Gln Tyr
        355                 360                 365
Ala Arg Ala Arg Lys Asp Phe Gln Leu Lys Ser His Asp Ser Val Phe
    370                 375                 380
Asp Val Lys Lys Gly Glu Leu Leu Cys Gly Tyr Gln Lys Val Val Met
385                 390                 395                 400
Thr Asp Pro Lys Val Phe Asp Glu Pro Glu Ser Phe Asn Ser Asp Arg
                405                 410                 415
Phe Val Gln Asn Ser Glu Leu Leu Asp Tyr Leu Tyr Trp Ser Asn Gly
            420                 425                 430
Pro Gln Thr Gly Thr Pro Thr Glu Ser Asn Lys Gln Cys Ala Ala Lys
        435                 440                 445
Asp Tyr Val Thr Leu Thr Ala Cys Leu Phe Val Ala Tyr Met Phe Arg
    450                 455                 460
Arg Tyr Asn Ser Val Thr Gly Ser Ser Ser Ser Ile Thr Ala Val Glu
465                 470                 475                 480
Lys Ala Asn
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Met Ala Arg Val Val Met Ser Asn
 1               5
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 488 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Ala Arg Val Val Met Ser Asn Met Ser Pro Ala Met Ser Ser Thr
 1               5                  10                  15
Tyr Pro Pro Ser Leu Ser Pro Pro Ser Ser Pro Arg Pro Thr Thr Leu
            20                  25                  30
Pro Val Arg Thr Ile Pro Gly Ser Tyr Gly Trp Pro Leu Leu Gly Pro
        35                  40                  45
Ile Ser Asp Arg Leu Asp Tyr Phe Trp Phe Gln Gly Pro Glu Thr Phe
    50                  55                  60
Phe Arg Lys Arg Ile Glu Lys Tyr Lys Ser Thr Val Phe Arg Ala Asn
65                  70                  75                  80
Val Pro Pro Cys Phe Pro Phe Phe Ser Asn Val Asn Pro Asn Val Val
                85                  90                  95
Val Val Leu Asp Cys Glu Ser Phe Ala His Leu Phe Asp Met Glu Ile
            100                 105                 110
Val Glu Lys Ser Asn Val Leu Val Gly Asp Phe Met Pro Ser Val Lys
        115                 120                 125
Tyr Thr Gly Asn Ile Arg Val Cys Ala Tyr Leu Asp Thr Ser Glu Pro
    130                 135                 140
Gln His Ala Gln Val Lys Asn Phe Ala Met Asp Ile Leu Lys Arg Ser
145                 150                 155                 160
Ser Lys Val Trp Glu Ser Glu Val Ile Ser Asn Leu Asp Thr Met Trp
                165                 170                 175
Asp Thr Ile Glu Ser Ser Leu Ala Lys Asp Gly Asn Ala Ser Val Ile
            180                 185                 190
Phe Pro Leu Gln Lys Phe Leu Phe Asn Phe Leu Ser Lys Ser Ile Ile
        195                 200                 205
Gly Ala Asp Pro Ala Ala Ser Pro Gln Val Ala Lys Ser Gly Tyr Ala
    210                 215                 220
Met Leu Asp Arg Trp Leu Ala Leu Gln Leu Leu Pro Thr Ile Asn Ile
225                 230                 235                 240
Gly Val Leu Gln Pro Leu Val Glu Ile Phe Leu His Ser Trp Ala Tyr
                245                 250                 255
Pro Phe Ala Leu Val Ser Gly Asp Tyr Asn Lys Leu Tyr Gln Phe Ile
            260                 265                 270
Glu Lys Glu Gly Arg Glu Ala Val Glu Arg Ala Lys Ala Glu Phe Gly
        275                 280                 285
Leu Thr His Gln Glu Ala Ile His Asn Leu Leu Phe Ile Leu Gly Phe
    290                 295                 300
Asn Ala Phe Gly Gly Phe Ser Ile Phe Leu Pro Thr Leu Leu Ser Asn
305                 310                 315                 320
Ile Leu Ser Asp Thr Thr Gly Leu Gln Asp Arg Leu Arg Lys Glu Val
                325                 330                 335
Arg Ala Lys Gly Gly Pro Ala Leu Ser Phe Ala Ser Val Lys Glu Met
            340                 345                 350
Glu Leu Val Lys Ser Val Val Tyr Glu Thr Leu Arg Leu Asn Pro Pro
        355                 360                 365
Val Pro Phe Gln Tyr Ala Arg Ala Arg Lys Asp Phe Gln Leu Lys Ser
    370                 375                 380
His Asp Ser Val Phe Asp Val Lys Lys Gly Glu Leu Leu Cys Gly Tyr
385                 390                 395                 400
Gln Lys Val Val Met Thr Asp Pro Lys Val Phe Asp Glu Pro Glu Ser
```

```
                405                 410                 415
Phe Asn Ser Asp Arg Phe Val Gln Asn Ser Glu Leu Leu Asp Tyr Leu
            420                 425                 430
Tyr Trp Ser Asn Gly Pro Gln Thr Gly Thr Pro Thr Glu Ser Asn Lys
        435                 440                 445
Gln Cys Ala Ala Lys Asp Tyr Val Thr Leu Thr Ala Cys Leu Phe Val
    450                 455                 460
Ala Tyr Met Phe Arg Arg Tyr Asn Ser Val Thr Gly Ser Ser Ser Ser
465                 470                 475                 480
Ile Thr Ala Val Glu Lys Ala Asn
                485
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1431 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
ATGTCGTCCA CCTACCCCCC GTCTCTGTCC CCGCCGTCGT CGCCGCGGCC GACCACCCTC     60
CCGGTGCGGA CGATCCCGGG CAGCTACGGG TGGCCCCTCC TCGGCCCGAT ATCGGACCGC    120
CTGGACTACT TCTGGTTCCA AGGCCCGGAG ACGTTCTTCA GGAAGAGGAT CGAGAAGTAC    180
AAGAGCACCG TGTTCCGCGC GAACGTGCCT CCGTGCTTCC CCTTCTTCTC GAACGTGAAC    240
CCTAACGTCG TGGTCGTCCT CGATTGCGAG TCCTTCGCTC ACTTGTTCGA CATGGAGATC    300
GTGGAGAAGA GCAACGTCCT CGTCGGCGAC TTCATGCCGA GCGTGAAGTA CACCGGGAAC    360
ATCCGGGTCT GCGCTTACCT CGACACTTCC GAGCCTCAAC ACGCTCAGGT GAAGAACTTT    420
GCGATGGACA TACTGAAGAG GAGCTCCAAA GTGTGGGAGA GCGAAGTGAT CTCGAACTTG    480
GACACCATGT GGGACACCAT CGAGTCCAGC CTCGCCAAGG ACGGCAACGC CAGCGTCATC    540
TTCCCTCTCC AAAAGTTCCT CTTCAACTTC CTCTCCAAGT CCATCATCGG CGCTGACCCG    600
GCCGCCTCGC CGCAGGTGGC CAAGTCCGGC TACGCCATGC TTGACCGGTG GCTCGCTCTC    660
CAGCTCCTCC CCACCATCAA CATTGGCGTA CTGCAGCCTC TAGTGGAGAT TTTTCTGCAT    720
TCTTGGGCAT ACCCTTTTGC GCTGGTGAGC GGGGACTACA ACAAGCTCTA CCAGTTCATC    780
GAGAAGGAAG GCCGAGAAGC GGTCGAAAGG GCGAAGGCCG AGTTCGGATT GACACACCAG    840
GAGGCCATCC ACAACTTGCT GTTCATCCTC GGCTTCAACG CGTTCGGCGG CTTCTCGATC    900
TTCCTCCCCA CGTTGCTGAG CAACATACTT AGCGACACAA CCGGACTGCA GGACCGGCTG    960
AGGAAGGAGG TCCGGGCAAA GGGAGGGCCG GCGTTGAGCT TCGCCTCGGT GAAGGAGATG   1020
GAACTCGTGA AGTCGGTCGT GTACGAGACG CTGCGGCTCA ACCCGCCCGT CCCGTTCCAA   1080
TACGCTCGAG CCCGGAAGGA CTTCCAGCTC AAGTCCCACG ACTCTGTCTT TGATGTCAAG   1140
AAAGGCGAGC TGCTATGCGG GTATCAGAAG GTGGTGATGA CAGACCCGAA AGTGTTCGAC   1200
GAACCGGAGA GCTTCAACTC GGACCGGTTC GTCCAAAACA GCGAGCTACT GGATTACCTG   1260
TACTGGTCCA ACGGGCCGCA GACCGGAACG CCGACCGAGT CGAACAAGCA GTGCGCGGCT   1320
AAGGACTACG TCACCCTCAC CGCTTGTCTC TTCGTTGCCT ACATGTTTCG ACGGTACAAT   1380
TCCGTCACAG GAAGCTCGAG CTCGATCACA GCCGTTGAAA AGGCCAACTG A            1431
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1443 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
ATGTCGCCGG CCATGTCGTC CACCTACCCC CCGTCTCTGT CCCCGCCGTC GTCGCCGCGG     60
CCGACCACCC TCCCGGTGCG GACGATCCCG GGCAGCTACG GGTGGCCCCT CCTCGGCCCG    120
ATATCGGACC GCCTGGACTA CTTCTGGTTC CAAGGCCCGG AGACGTTCTT CAGGAAGAGG    180
ATCGAGAAGT ACAAGAGCAC CGTGTTCCGC GCGAACGTGC CTCCGTGCTT CCCCTTCTTC    240
TCGAACGTGA ACCCTAACGT CGTGGTCGTC CTCGATTGCG AGTCCTTCGC TCACTTGTTC    300
GACATGGAGA TCGTGGAGAA GAGCAACGTC CTCGTCGGCG ACTTCATGCC GAGCGTGAAG    360
TACACCGGGA ACATCCGGGT CTGCGCTTAC CTCGACACTT CCGAGCCTCA ACACGCTCAG    420
GTGAAGAACT TTGCGATGGA CATACTGAAG AGGAGCTCCA AAGTGTGGGA GAGCGAAGTG    480
ATCTCGAACT TGGACACCAT GTGGGACACC ATCGAGTCCA GCCTCGCCAA GGACGGCAAC    540
GCCAGCGTCA TCTTCCCTCT CCAAAAGTTC CTCTTCAACT TCCTCTCCAA GTCCATCATC    600
GGCGCTGACC CGGCCGCCTC GCCGCAGGTG GCCAAGTCCG GCTACGCCAT GCTTGACCGG    660
TGGCTCGCTC TCCAGCTCCT CCCCACCATC AACATTGGCG TACTGCAGCC TCTAGTGGAG    720
ATTTTTCTGC ATTCTTGGGC ATACCCTTTT GCGCTGGTGA GCGGGGACTA CAACAAGCTC    780
TACCAGTTCA TCGAGAAGGA AGGCCGAGAA GCGGTCGAAA GGGCGAAGGC CGAGTTCGGA    840
TTGACACACC AGGAGGCCAT CCACAACTTG CTGTTCATCC TCGGCTTCAA CGCGTTCGGC    900
GGCTTCTCGA TCTTCCTCCC CACGTTGCTG AGCAACATAC TTAGCGACAC AACCGGACTG    960
CAGGACCGGC TGAGGAAGGA GGTCCGGGCA AAGGGAGGGC CGGCGTTGAG CTTCGCCTCG   1020
GTGAAGGAGA TGGAACTCGT GAAGTCGGTC GTGTACGAGA CGCTGCGGCT CAACCCGCCC   1080
GTCCCGTTCC AATACGCTCG AGCCCGGAAG GACTTCCAGC TCAAGTCCCA CGACTCTGTC   1140
TTTGATGTCA AGAAAGGCGA GCTGCTATGC GGGTATCAGA AGGTGGTGAT GACAGACCCG   1200
AAAGTGTTCG ACGAACCGGA GAGCTTCAAC TCGGACCGGT TCGTCCAAAA CAGCGAGCTA   1260
CTGGATTACC TGTACTGGTC CAACGGGCCG CAGACCGGAA CGCCGACCGA GTCGAACAAG   1320
CAGTGCGCGG CTAAGGACTA CGTCACCCTC ACCGCTTGTC TCTTCGTTGC CTACATGTTT   1380
CGACGGTACA ATTCCGTCAC AGGAAGCTCG AGCTCGATCA CAGCCGTTGA AAAGGCCAAC   1440
TGA                                                                 1443
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 1452 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
ATGAGCAACA TGTCGCCGGC CATGTCGTCC ACCTACCCCC CGTCTCTGTC CCCGCCGTCG     60
TCGCCGCGGC CGACCACCCT CCCGGTGCGG ACGATCCCGG GCAGCTACGG GTGGCCCCTC    120
CTCGGCCCGA TATCGGACCG CCTGGACTAC TTCTGGTTCC AAGGCCCGGA GACGTTCTTC    180
AGGAAGAGGA TCGAGAAGTA CAAGAGCACC GTGTTCCGCG CGAACGTGCC TCCGTGCTTC    240
CCCTTCTTCT CGAACGTGAA CCCTAACGTC GTGGTCGTCC TCGATTGCGA GTCCTTCGCT    300
CACTTGTTCG ACATGGAGAT CGTGGAGAAG AGCAACGTCC TCGTCGGCGA CTTCATGCCG    360
AGCGTGAAGT ACACCGGGAA CATCCGGGTC TGCGCTTACC TCGACACTTC CGAGCCTCAA    420
CACGCTCAGG TGAAGAACTT TGCGATGGAC ATACTGAAGA GGAGCTCCAA AGTGTGGGAG    480
AGCGAAGTGA TCTCGAACTT GGACACCATG TGGGACACCA TCGAGTCCAG CCTCGCCAAG    540
GACGGCAACG CCAGCGTCAT CTTCCCTCTC CAAAAGTTCC TCTTCAACTT CCTCTCCAAG    600
TCCATCATCG GCGCTGACCC GGCCGCCTCG CCGCAGGTGG CCAAGTCCGG CTACGCCATG    660
CTTGACCGGT GGCTCGCTCT CCAGCTCCTC CCCACCATCA ACATTGGCGT ACTGCAGCCT    720
CTAGTGGAGA TTTTTCTGCA TTCTTGGGCA TACCCTTTTG CGCTGGTGAG CGGGGACTAC    780
AACAAGCTCT ACCAGTTCAT CGAGAAGGAA GGCCGAGAAG CGGTCGAAAG GGCGAAGGCC    840
GAGTTCGGAT TGACACACCA GGAGGCCATC CACAACTTGC TGTTCATCCT CGGCTTCAAC    900
GCGTTCGGCG GCTTCTCGAT CTTCCTCCCC ACGTTGCTGA GCAACATACT TAGCGACACA    960
ACCGGACTGC AGGACCGGCT GAGGAAGGAG GTCCGGGCAA AGGGAGGGCC GGCGTTGAGC   1020
TTCGCCTCGG TGAAGGAGAT GGAACTCGTG AAGTCGGTCG TGTACGAGAC GCTGCGGCTC   1080
AACCCGCCCG TCCCGTTCCA ATACGCTCGA GCCCGGAAGG ACTTCCAGCT CAAGTCCCAC   1140
GACTCTGTCT TTGATGTCAA GAAAGGCGAG CTGCTATGCG GGTATCAGAA GGTGGTGATG   1200
ACAGACCCGA AAGTGTTCGA CGAACCGGAG AGCTTCAACT CGGACCGGTT CGTCCAAAAC   1260
AGCGAGCTAC TGGATTACCT GTACTGGTCC AACGGGCCGC AGACCGGAAC GCCGACCGAG   1320
TCGAACAAGC AGTGCGCGGC TAAGGACTAC GTCACCCTCA CCGCTTGTCT CTTCGTTGCC   1380
```

TACATGTTTC GACGGTACAA TTCCGTCACA GGAAGCTCGA GCTCGATCAC AGCCGTTGAA 1440
AAGGCCAACT GA 1452

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1467 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATGGCGAGGG TCGTGATGAG CAACATGTCG CCGGCCATGT CGTCCACCTA CCCCCCGTCT 60
CTGTCCCCGC CGTCGTCGCC GCGGCCGACC ACCCTCCCGG TGCGGACGAT CCCGGGCAGC 120
TACGGGTGGC CCCTCCTCGG CCCGATATCG GACCGCCTGG ACTACTTCTG GTTCCAAGGC 180
CCGGAGACGT TCTTCAGGAA GAGGATCGAG AAGTACAAGA GCACCGTGTT CCGCGCGAAC 240
GTGCCTCCGT GCTTCCCCTT CTTCTCGAAC GTGAACCCTA ACGTCGTGGT CGTCCTCGAT 300
TGCGAGTCCT TCGCTCACTT GTTCGACATG GAGATCGTGG AGAAGAGCAA CGTCCTCGTC 360
GGCGACTTCA TGCCGAGCGT GAAGTACACC GGGAACATCC GGGTCTGCGC TTACCTCGAC 420
ACTTCCGAGC CTCAACACGC TCAGGTGAAG AACTTTGCGA TGGACATACT GAAGAGGAGC 480
TCCAAAGTGT GGGAGAGCGA AGTGATCTCG AACTTGGACA CCATGTGGGA CACCATCGAG 540
TCCAGCCTCG CCAAGGACGG CAACGCCAGC GTCATCTTCC CTCTCCAAAA GTTCCTCTTC 600
AACTTCCTCT CCAAGTCCAT CATCGGCGCT GACCCGGCCG CCTCGCCGCA GGTGGCCAAG 660
TCCGGCTACG CCATGCTTGA CCGGTGGCTC GCTCTCCAGC TCCTCCCCAC CATCAACATT 720
GGCGTACTGC AGCCTCTAGT GGAGATTTTT CTGCATTCTT GGGCATACCC TTTTGCGCTG 780
GTGAGCGGGG ACTACAACAA GCTCTACCAG TTCATCGAGA AGGAAGGCCG AGAAGCGGTC 840
GAAAGGGCGA AGGCCGAGTT CGGATTGACA CACCAGGAGG CCATCCACAA CTTGCTGTTC 900
ATCCTCGGCT TCAACGCGTT CGGCGGCTTC TCGATCTTCC TCCCCACGTT GCTGAGCAAC 960
ATACTTAGCG ACACAACCGG ACTGCAGGAC CGGCTGAGGA AGGAGGTCCG GGCAAAGGGA 1020
GGGCCGGCGT TGAGCTTCGC CTCGGTGAAG GAGATGGAAC TCGTGAAGTC GGTCGTGTAC 1080
GAGACGCTGC GGCTCAACCC GCCCGTCCCG TTCCAATACG CTCGAGCCCG GAAGGACTTC 1140
CAGCTCAAGT CCCACGACTC TGTCTTTGAT GTCAAGAAAG GCGAGCTGCT ATGCGGGTAT 1200
CAGAAGGTGG TGATGACAGA CCCGAAAGTG TTCGACGAAC CGGAGAGCTT CAACTCGGAC 1260
CGGTTCGTCC AAAACAGCGA GCTACTGGAT TACCTGTACT GGTCCAACGG GCCGCAGACC 1320
GGAACGCCGA CCGAGTCGAA CAAGCAGTGC GCGGCTAAGG ACTACGTCAC CCTCACCGCT 1380
TGTCTCTTCG TTGCCTACAT GTTTCGACGG TACAATTCCG TCACAGGAAG CTCGAGCTCG 1440
ATCACAGCCG TTGAAAAGGC CAACTGA 1467

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1443 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATGATACCTA TAATGAGCTC TGCTCCTCTA TCAACTGCTA CACCAATATC TCTCCCCGTA 60
CGTAAAATTC CAGGGAGCTA CGGGTTTCCA TTATTAGGGC CACTTTGGGA TCGATTAGAC 120
TATAACTGGT TCCAAAAGCT CCCAGATTTC TTCAGCAAGA GAGTCGAAAA GTATAACAGC 180
ACGGTATTCA GAACGAATGT ACCGCCTTGT TTTCCATTTT TTTTGGGTGT AAATCCAAAT 240
GTAGTGGCGG TACTGGATGT CAAGTCATTT GCACATCTAT TTGATATGGA GATTGTTGAG 300
AAAGCTAATG TGCTTGTTGG TGATTTCATG CCCAGTGTTG TTTATACTGG TGATATGCGT 360
GTTTGTGCTT ATCTTGATAC TTCTGAACCT AAACATACTC AGATTAAGAA CTTTTCATTG 420
GACATCCTAA AAAGAAGTTC AAAGACATGG GTGCCTACAC TAGTTAAAGA ACTTGATACA 480

```
CTGTTCGGAA CTTTTGAATC AGATCTTTCA AAATCCAAAT CAGCTTCTCT TCTCCCTGCA    540
TTGCAAAAAT TCCTCTTCAA CTTCTTCTCC TTAACTTTCC TCGGGGCCGA TCCATCAGCC    600
TCACCGGAGA TAGCCAACTC TGGCTTCGCC TATCTTGATG CATGGCTAGC TATTCAACTA    660
GCACCTACTG TTAGCATTGG TGTTCTTCAA CCCCTTGAAG AAATCTTCGT CCACTCTTTT    720
TCATACCCCT ATTTTCTTGT CCGTGGAGGT TACGAAAAAC TCATTAAGTT TGTGAAAAGT    780
GAAGCTAAGG AAGTGTTAAC GAGGGCACAA ACAGACTTTC AGCTAACTGA ACAAGAAGCC    840
ATTCATAACC TTTTGTTCAT TCTTGGATTC AATGCTTTTG GTGGTTTCAC CATTTTCTTG    900
CCAACCCTTC TGGGAAACCT TGGGAGACGA GAAAAATGCT GAGATGCAAG AGAAACTGAG    960
AAAAGAAGTG AGGGAAAAAG TTGGACAAAT CAAGAAAACT TGAGTTTTGA GAGTGTAAAA   1020
GAAATGGAAC TGGTTCAGTC TTTTGTTTAT GAATCACTTA GGCTAAGCCC ACCAGTGCCA   1080
AGTCAATATG CAAGAGCAAG AAAAGACTTC ATGCTCAGTT CACATGATTC AGTTTACGAA   1140
ATCAAGAAAG GTGAACTTCT TTGTGGTTAC CAGCCATTAG TGATGAAAGA TCCAAAGGTG   1200
TTTGATGAAC CTGAAAAGTT TATGTTGGAG AGGTTTACAA AGGAGAAAGG GAAAGAATTG   1260
CTGAATTATT TGTTTTGGTC TAATGGCCCA CAGACTGGGA GCCCTACTGA ATCAAACAAG   1320
CAATGTGCTG CTAAGGATGC GGTTACTCTT ACTGCTTCTT TGATTGTGGC TTACATTTTC   1380
CAAAAGTATG ATTCTGTGAG TTTCTCATCT GGTTCACTCA CATCTGTGAA AAAAGCCTGC   1440
TGA                                                                1443
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 1638 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
AAGAAGAAGA GAGGGAAGGT ACGGATGGCT ATGATGTGGT CGTCAGCCTC CGCCACCGCC     60
GTCACCACGC TGCCGACGAG GCCCATCCCT GGAAGCTACG GCCCGCCGCT GGTGGGCCCC    120
CTCAAGGACC GCCTCGACTA CTTCTGGTTT CAGGGACCGG AGACCTTCTT CCGCAGCCGG    180
ATGGCCACCC ACAAGAGCAC CGTGTTCCGC ACCAACATGC CCCCCACCTT CCCCTTCTTC    240
GTTGGAGTCG ACCCCCGCGT GGTCACCGTC CTCGACTGCA CATCCTTCTC CGCCCTCTTC    300
GACCTCGAGG TCGTGGAGAA GAAGAACATT CTCATCGGGG ACTACATGCC CAGCCTCAGC    360
TTCACCGGCG ACACCCGCGT CGTCGTGTAC CTCGACCCCT CCGAGCCCGA CCACGCCCGC    420
GTGAAGAGCT TCTGCTTGGA ACTCCTCAGG CGCGGCGCCA AGACCTGGGT CTCCTCGTTC    480
CTCTCCAATC TCGATGTCAT GCTCGCCACC ATAGAGCAGG GGATCGCCAA GGATGGCTCC    540
GCCGGCTTAT TCGGCCCGCT GCAGAAGTGC ATCTTCGCGT TCCTCTGCAA GAGCATCATC    600
GGGGCCGACC CGTCGGTGTC GCCCGACGTG GGAGAAAATG GCTTCGTCAT GCTCGACAAG    660
TGGCTTGCGC TGCAGCTCCT CCCGACGGTG AAGGTCGGGG CCATCCCGCA ACCCCTGGAG    720
GAGATCCTCC TCCACTCCTT CCCCCTCCCC TTCTTCCTCG TGAGCCGCGA TTACCGGAAG    780
CTGTACGAAT TCGTCGAGAA GCAAGGCCAA GAGGTTGTCC GGCGAGCGGA AACCGAGCAC    840
GGGCTCAGCA AGCACGACGC CATCAACAAC ATCTTGTTCG TCCTAGGATT CAACGCCTTC    900
GGCGGCTTCT CGGTCTTCTT CCCCACGCTC CTGACCACCA TAGGGAGGGA CAAGACGGGC    960
CTGCGGGAGA AGCTCAAGGA CGAGGTGCGC AGGGTCATGA AGAGTAGAGG GGAGAAGCGG   1020
CCGAGCTTCG AGACGGTGCG GGAGATGGAG CTGGTGCGAT CGACGGTGTA CGAGGTCCTG   1080
CGGCTGAACC CGCCGGTGCC GCTGCAGTAC GGGCGGGCGC GCACCGACTT CACGCTGAAC   1140
TCCCACGACG CGGCGTTCAA GGTTGAGAAG GGGGAGTTGC TGTGCGGGTA CCAGCCGCTG   1200
GTGATGCGGG ATCCAGCGGT GTTCGACGAC CCGGAGACGT TCGCCCCGGA AAGGTTCATG   1260
GGCAGCGGGA AGGAGCTGCT CAAGTACGTC TTCTGGTCCA ACGGGCCGGA GACGGGTACG   1320
CCGACGCCGG CCAACAAGCA GTGCGCCGCG AAGGACTACG TGGTGGAGAC GGCGTGCCTG   1380
CTGATGGCGG AGATCTTCTA CCGCTACGAC GAGTTCGTGT GCGCCGACGA CGCCATCTCC   1440
GTGACGAAGC TGGATAGAGC GAGAGAATGG GAGTAAACGG TATTCAAGTC GGAAGCGACA   1500
TAAGGAGACG GCCAACTCCA CCGTTGCTAA TTCAAGTCGT ACTCCAAATC GGTATTCATA   1560
TCATCGTTCC ATTGGGGTGA TGAAGAGATA AATAAAATTT GACGTTGCAG GAGGCTACAA   1620
AAAAAAAAAA AAAAAAAA                                                 1638
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
Asp Gly Asn Ala Ser Val Ile Phe Pro Leu Gln
 1               5                  10
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Asn Phe Ala Met Asp Ile Leu
 1               5
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
Phe Leu Phe Asn Phe Leu Ser
 1               5
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
GCGGATCCGG CCATGAGCAA CATGTCG                                    27
```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AATGTTGATG GTGGGGAGGA G 21

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GCGGATCCGG CCATGTCGCC GGCCAT 26

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GCGGATCCGG CCATGTCGTC CACCTAC 27

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Thr Tyr Pro Pro Ser Leu Ser Pro
1 5

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Thr Tyr Pro Pro Ser Leu Ser Pro Pro Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Thr Tyr Pro Pro Ser Leu Ser Pro Pro Ser Ser Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Thr Tyr Pro Pro Ser Leu Ser Pro Pro Ser Ser Pro Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Thr Tyr Pro Pro Ser Leu Ser Pro Pro Ser Ser Pro Arg Pro
1 5 10

**What is claimed is:**

1. An isolated protein comprising a fatty acid 13-hydroperoxide lyase comprising the amino acid sequence set forth in SEQ ID NO:1.

2. The protein of claim 1, wherein the amino acid sequence of the fatty acid 13-hydroperoxide lyase is an amino acid sequence present in fatty acid 13-hydroperoxide lyase isolated from *Psidium guajava*.

3. An isolated protein comprising the amino acid sequence set forth in SEQ ID NO:2.

4. An isolated protein comprising the amino acid sequence set forth in SEQ ID NO:3.

5. An isolated protein comprising the amino acid sequence set forth in SEQ ID NO:4.

6. An isolated protein comprising a fatty acid 13-hydroperoxide lyase comprising at its N-terminus the amino acid sequence set forth in SEQ ID NO:5.

7. The protein of claim 6, wherein the amino acid sequence of the fatty acid 13-hydroperoxide lyase is an amino acid sequence present in fatty acid 13-hydroperoxide lyase isolated from *Psidium guajava*.

8. An isolated protein comprising a polypeptide having the amino acid sequence set forth in SEQ ID NO:6.

9. An isolated nucleic acid comprising a nucleic acid encoding the protein of claim 1.

10. An isolated nucleic acid comprising a nucleic acid encoding the protein of claim 3.

11. An isolated nucleic acid comprising a nucleic acid encoding the protein of claim 4.

12. An isolated nucleic acid comprising a nucleic acid encoding the protein of claim 5.

13. An isolated nucleic acid comprising a nucleic acid encoding the protein of claim 6.

14. An isolated nucleic acid comprising a nucleic acid encoding the protein of claim 8.

15. The nucleic acid of claim 9, wherein the nucleic acid has the nucleotide sequence set forth in SEQ ID NO:7.

16. The nucleic acid of claim 11, wherein the nucleic acid has the nucleotide sequence set forth in SEQ ID NO:8.

17. The nucleic acid of claim 12, wherein the nucleic acid has the nucleotide sequence set forth in SEQ ID NO:9.

18. The nucleic acid of claim 13, wherein the nucleic acid has the nucleotide sequence set forth in SEQ ID NO:10.

19. An isolated nucleic acid which specifically hybridizes with the nucleic acid of SEQ ID NO:7 under stringent conditions of hybridization and which does not hybridize at the stringent conditions to the nucleic acid set forth in SEQ ID NO:11 or SEQ ID NO:12.

20. The isolated nucleic acid of claim 19 wherein the nucleic acid has at least 90% complementarity with the sequence to which it hybridizes.

21. The isolated nucleic acid of claim 19 wherein the nucleic acid has at least 80% complementarity with the sequence to which it hybridizes.

22. The isolated nucleic acid of claim 19 wherein the nucleic acid has at least 70% complementarity with the sequence to which it hybridizes.

23. The isolated nucleic acid of claim 19 wherein the nucleic acid encodes a functional 13-hydroperoxide lyase.

24. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising the nucleic acid of claim 9.

25. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising the nucleic acid of claim 10.

26. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising the nucleic acid of claim 11.

27. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising the nucleic acid of claim 12.

28. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising the nucleic acid of claim 13.

29. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising the nucleic acid of claim 14.

30. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising the nucleic acid of claim 15.

31. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising the nucleic acid of claim 16.

32. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising the nucleic acid of claim 17.

33. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising the nucleic acid of claim 18.

34. The vector of claim 24, wherein the vector is a plasmid.

35. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising a promoter functionally linked to the nucleic acid of claim 9.

36. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising a promoter functionally linked to the nucleic acid of claim 10.

37. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising a promoter functionally linked to the nucleic acid of claim 11.

38. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising a promoter functionally linked to the nucleic acid of claim 12.

39. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising a promoter functionally linked to the nucleic acid of claim 13.

40. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising a promoter functionally linked to the nucleic acid of claim 14.

41. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising a promoter functionally linked to the nucleic acid of claim 15.

42. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising a promoter functionally linked to the nucleic acid of claim 16.

43. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising a promoter functionally linked to the nucleic acid of claim 17.

44. A vector for the expression of a *Psidium guajava* 13-hydroperoxide lyase comprising a promoter functionally linked to the nucleic acid of claim 18.

45. A cell containing an exogenous nucleic acid comprising the nucleic acid of claim 9.

46. The cell of claim 45, wherein the cell is an *Escherichia coli* cell.

47. The cell of claim 45, wherein the cell is a yeast cell.

48. A method of cleaving a 13-hydroperoxide of linoleic or α-linolenic acid into a $C_6$- aldehyde and a $C_{12}$-oxocarboxylic acid comprising contacting the

recombinant protein produced by the vector of claim 24 with the 13-hydroperoxide, thereby cleaving the 13-hydroperoxide.

49. A method of preparing n-hexanal, 3-(Z)-hexen-1-al, 2-(E)-hexen-1-al, or their corresponding alcohols from 13-hydroperoxy-octadeca-9,11-dienoic acid or 13 hydroperoxy-octadeca-9,11,15-trienoic acid, comprising

(a) contacting the 13-hydroperoxy-octadeca-9,11-dienoic acid or the 13 hydroperoxy-octadeca-9,11,15-trienoic acid with the recombinant protein produced by the vector of claim 24, thereby converting the 13-hydroperoxy-octadeca-9,11-dienoic acid into n-hexanal or the 13 hydroperoxy-octadeca-9,11,15-trienoic acid into 3-(Z)-hexen-1-al; and either

(b) recovering the n-hexanal or 3-(Z)-hexen-1-al;

(b′) reducing the n-hexanal into n-hexanol or the 3-(Z)-hexen-1-al into 3-(Z)-hexen-1-ol and recovering the hexanol or 3-(Z)-hexen-1-ol; or

(b′′) isomerizing the 3-(Z)-hexen-1-al under temperature and pH conditions effective to obtain 2-(E)-hexen-1-al and either recovering the formed 2-(E)-hexen-1-al or reducing the 2-(E)-hexen-1-al to 2-(E)-hexen-1-ol and recovering the 2-(E)-hexen-1-ol from the medium.

50. A method of expressing a recombinant protein produced by the cell of claim 45, comprising optimizing active lyase function of the recombinant protein by culturing the cells in the absence of isopropyl β-D-thiogalactopyranoside.

51. The method of claim 50, further comprising culturing the cells in the absence of δ-aminolevulinic acid.

## ABSTRACT

The present invention relates to fatty acid 13-hydroperoxide lyase protein from guava (*Psidium guajava)* and the gene encoding the protein. Expression systems for recombinant guava 13-hydroperoxide lyase and methods of using recombinant guava 13-hydroperoxide lyase for the production of green notes are provided.

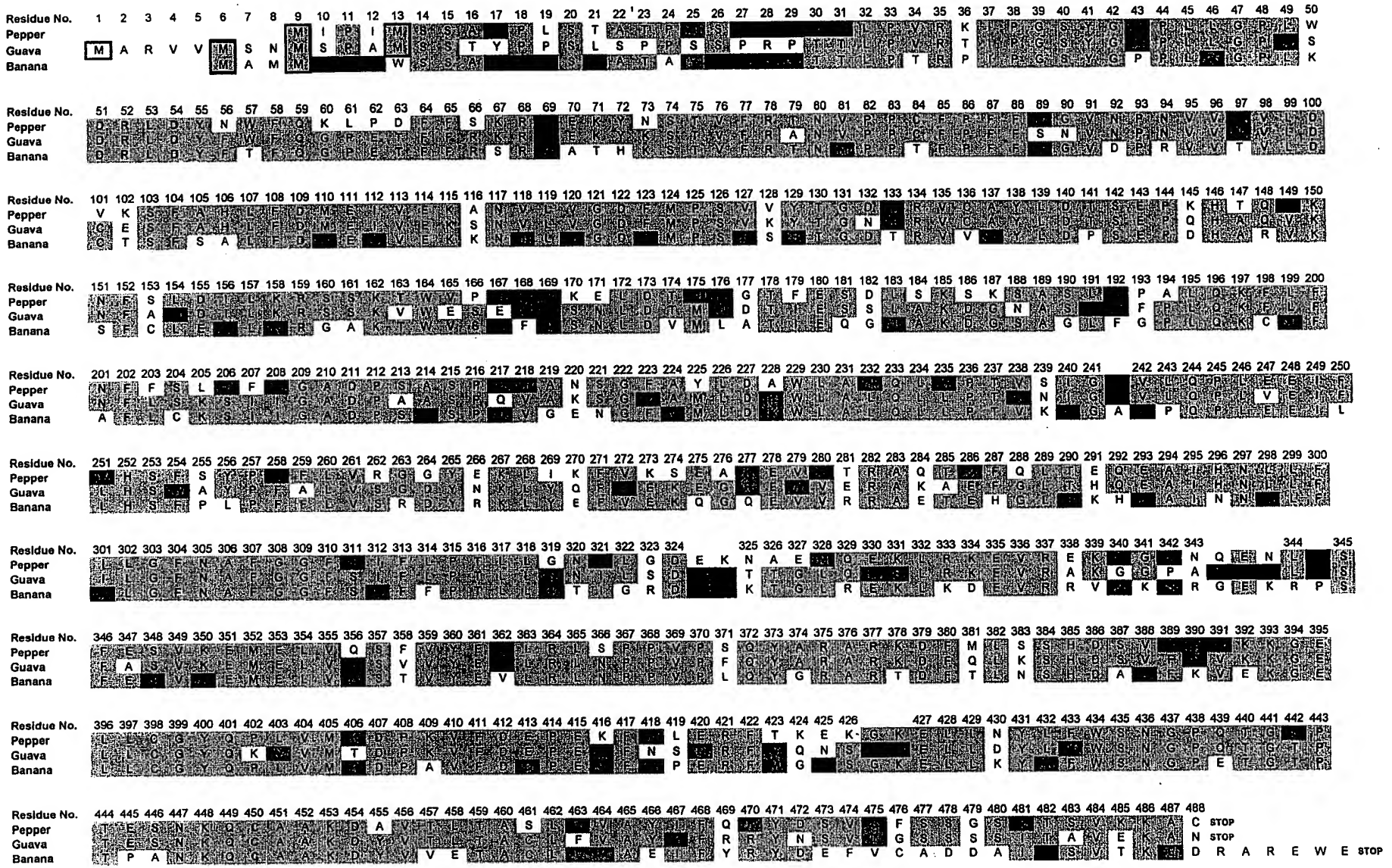


FIG. 1

"GUAVA (PSIDIUM GUAJAVA)
13-HYDROPEROXIDE LYASE AND USES
THEREOF" (*Whitehead et al.*)

# Complete Nucleotide and Derived Amino Acid Sequence of Guava 13-HPOL

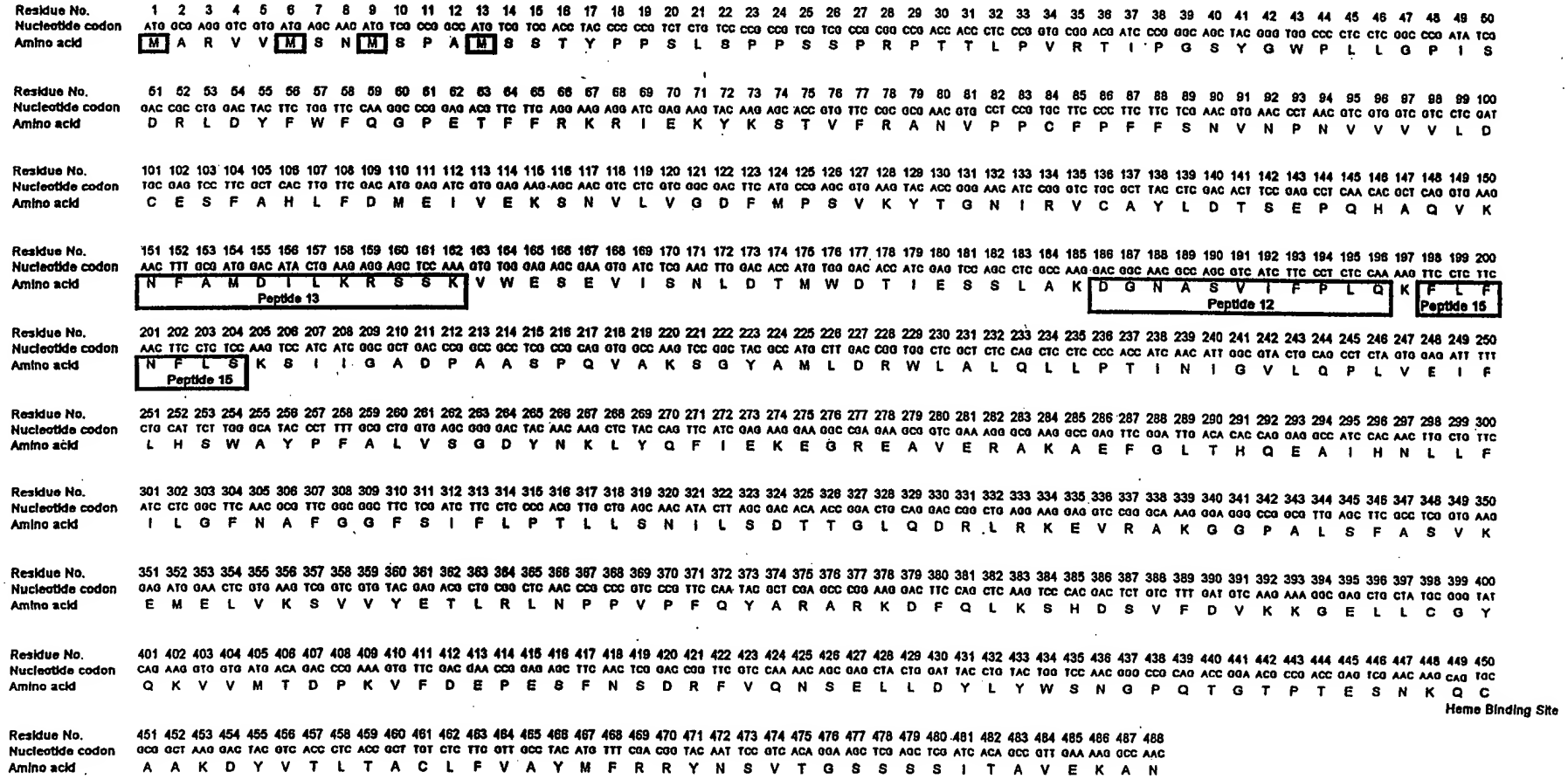


FIG. 2